18

Phases of Gauge Theories

So far we have primarily focused on the behavior of quantum field theories within perturbation theory in powers of a weak coupling constant g. However, we have also seen that perturbation theory often fails in the sense that, even in the perturbative regime, the effective (or renormalized) coupling constant runs to string coupling at low-energies (the IR). We should regard this behavior as an indication that the fixed point at $g = 0$ represents an *unstable* ground state , and that the true ground state has very different physics.

In some cases we were able to use the the $1/N$ expansion to investigate the strong coupling regime. Unfortunately, the the large N_c limit of gauge theories (here N_c is the number of colors) is not solvable. It has been believed since the late 1970's that their ground states represent a phase in which the theories are *confining*. As we noted in Chapter 17, the $N_c \to \infty$ limit of a supersymmetric version of Yang-Mills theory was solved using the gauge/gravity duality, the Maldacena conjecture. Even in the case, confinement could only be proven by breaking the supersymmetry (which breaks the conformal invariance). Although much progress has been made, the existing proof of confinement remain qualitative in character.

In this chapter we will discuss a non-perturbative approach to quantum field theory in which the theory is regularized in the UV by placing the degrees of freedom on some (typically hypercubic) lattice.

18.1 Lattice regularization of QFT

The strong coupling behavior of essentially any field theory can be studied by defining the theory with a lattice UV cutoff. In this representation, pioneered by K. G. Wilson and others, most theories (including gauge theories) can be studied by using methods borrowed from Statistical Mechanics.

In this framework the theory is defined in discretized D-dimensional Euclidean space-time, i.e. a D-dimensional hypercubic lattice of lattice spacing a_0 (which, most of the time, we will set to 1). The path-integral of the Euclidean quantum field theory becomes a partition function in classical statistical mechanics on a hypercubic lattice. In this version of the theory, one of the lattice directions is interpreted as a discretized imaginary time while the $D-1$ remaining directions correspond to the spatial directions. While in the lattice theories rotational invariance (the Euclidean version of Lorentz invariance) is broken by the lattice, but the the discrete point group symmetries of the hypercubic lattice are preserved.However, rotational invariance is recovered at a non-trivial fixed point, at which the correlation length diverges.

In this perspective, since in the path integral the Planck constant $\hbar$ enters in the same way as the temperature T in classical statistical mechanics, weak coupling perturbation theory is equivalent to the low temperature expansion in classical statistical mechanics. Conversely, the strong coupling regime of the quantum field theory is identified with the high temperature regime of statistical mechanics, which is described by the high temperature expansion.

At least in principle, this connection allows one to access the string coupling regime but at the price of working with a cutoff. As we saw in previous chapters, the cutoff can only be removed by investigating the behavior of the theory close to a non-trivial fixed point. But to do this it is necessary to extrapolate these expansions outside their radii of convergence! Another virtue of this approach is that, at least in abelian theories, it is possible to make duality transformations that relate two (generally different) theories one at weak coupling and the other at strong coupling!

Finally, once these theories are defined on a lattice it is possible to use powerful numerical techniques, such as Monte Carlo simulations, to investigate their properties and phase transitions. This numerical approach has become a powerful tool to compute the low-energy spectrum of hadrons, as well as for the study of thermodynamic properties of gauge theory at finite temperature.

18.2 Matter Fields

Let $\{\boldsymbol{r}\}$ be the coordinates of the sites of this lattice and, hence, $\boldsymbol{r}$ represents a D-tuplet of integer-valued numbers. Let $\mu = 1, \ldots, D$ label the D possible directions on the hypercubic lattice. A link of the lattice are thus labeled by the coordinates of a site and a direction, $(\boldsymbol{r}, \mu)$. Thus, if $\boldsymbol{r}$ is a lattice site, $\boldsymbol{r} + \boldsymbol{e}_\mu$ is the nearest-neighbor site along the direction of the unit vector

$\boldsymbol{e}_\mu$ (with $\mu = 1, \ldots, D$). At case in $D = 2$ dimensions is shown in Fig.18.1. The matter fields are defined to be on the lattice sites. Let us consider first

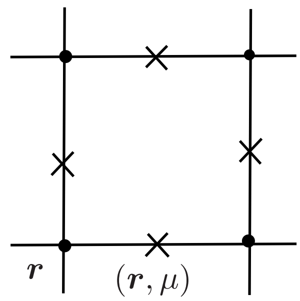

Figure 18.1 A two-dimensional euclidean lattice. Here $\boldsymbol{r}$ are the lattice sites and $(\boldsymbol{r}, \mu)$ are the links.

the case in which the matter field is a scalar field. Let $\phi(\boldsymbol{r})$ be a scalar (matter) field, transforming in some irreducible representation of a compact Lie group G. For instance, in the case of a principal chiral field, the field $\phi(\boldsymbol{r})$ is a group element, and obeys the global transformation law,

$$\phi'(\boldsymbol{r}) = V\phi(\boldsymbol{r})U^{-1} \tag{18.1}$$

The lattice action must be invariant under global transformations. Thus, it must be written in terms of G-invariant quantities. A typical term is

$$\operatorname{tr}\sum_{\boldsymbol{r}} \left(\phi^\dagger(\boldsymbol{r})\phi(\boldsymbol{r} + \boldsymbol{e}_\mu)\right) \tag{18.2}$$

which is clearly invariant under the global symmetry:

$$\begin{aligned}\operatorname{tr}\sum_{\boldsymbol{r}} \left(\phi'^\dagger(\boldsymbol{r})\phi'(\boldsymbol{r} + \boldsymbol{e}_\mu)\right) =& \operatorname{tr}\sum_{\boldsymbol{r}} \left(U\phi'^\dagger(\boldsymbol{r})V^{-1}V\phi'(\boldsymbol{r} + \boldsymbol{e}_\mu)U^{-1}\right)\\ =& \operatorname{tr}\sum_{\boldsymbol{r}} \left(\phi^\dagger(\boldsymbol{r})\phi(\boldsymbol{r} + \boldsymbol{e}_\mu)\right)\end{aligned} \tag{18.3}$$

since $U^\dagger = U^{-1}$ and where we use the cyclic invariance of the trace. We will consider field theories in which the group is a) $G = \mathbb{Z}_2$, in which case the matter fields are Ising spins, $\sigma(\boldsymbol{r}) = \pm 1$, b) $G = U(1)$, in which case the matter fields are phases (i.e. XY spins) , $\exp(i\theta(\boldsymbol{r}))$, with $\theta(\boldsymbol{r}) \in [0, 2\pi)$, c) $G = O(3)$ and the degrees of freedom are unit-length real vectors, $\boldsymbol{n}(\boldsymbol{r})$ (such that $\boldsymbol{n}^2(\boldsymbol{r}) = 1$) (i.e. the non-linear sigma model), and so on.

18.3 Minimal Coupling

In order to promote the global symmetry G to a local symmetry we need to introduce gauge fields and minimal coupling. The gauge fields are connections (vector fields) and are defined to be on the links of the lattice. In general the gauge fields will be group elements $\mathcal{U}_\mu(\boldsymbol{r}) \in G$ defined on the links $(\boldsymbol{r}, \mu)$ of the lattice, such that $\mathcal{U}_\mu(\boldsymbol{r}) = \mathcal{U}_\mu^{-1}(\boldsymbol{r} + \mathrm{e}_\mu)$. Since they are group elements they can be written as

$$\mathcal{U}_\mu(\boldsymbol{r}) = \exp(iA_\mu(\boldsymbol{r})) \tag{18.4}$$

where $A_\mu(\boldsymbol{r})$ is an element in the algebra of the Lie group G, i.e. $A_\mu(\boldsymbol{r}) = A_\mu^k(\boldsymbol{r})\lambda^k$, where λ^k are the generators of G. Clearly $A_\mu(\boldsymbol{r})$ can be interpreted as implementing parallel transport between two lattice sites. Hence

$$A_\mu(\boldsymbol{r}) \sim \int_{\boldsymbol{r}}^{\boldsymbol{r}+\boldsymbol{e}_\mu} dx_\nu A^\nu(x) \sim a_0 A_\mu(\boldsymbol{r}) \tag{18.5}$$

A local gauge transformation is

$$\mathcal{U}_\mu'(\boldsymbol{r}) = V(\boldsymbol{r})\mathcal{U}_\mu(\boldsymbol{r})V^{-1}(\boldsymbol{r} + \boldsymbol{e}_\mu), \quad \phi'(\boldsymbol{r}) = V(\boldsymbol{r})\phi(\boldsymbol{r}) \tag{18.6}$$

A gauge-invariant term in the action involves the lattice version of a covariant derivative

$$\begin{aligned} \mathrm{tr}\left[\phi'^\dagger(\boldsymbol{r})\mathcal{U}_\mu'(\boldsymbol{r})\phi'(\boldsymbol{r} + \boldsymbol{e}_\mu)\right] =& \mathrm{tr}\left[\phi^\dagger(\boldsymbol{r})V^{-1}(\boldsymbol{r})\mathcal{U}_\mu'(\boldsymbol{r})V(\boldsymbol{r} + \boldsymbol{e}_\mu)\phi(\boldsymbol{r} + \boldsymbol{e}_\mu)\right] \\ =& \mathrm{tr}\left[\phi^\dagger(\boldsymbol{r})\mathcal{U}_\mu(\boldsymbol{r})\phi(\boldsymbol{r} + \boldsymbol{e}_\mu)\right] \end{aligned} \tag{18.7}$$

The Euclidean action for the matter field is

$$-S_{\text{matter}} = \frac{\beta}{2}\sum_{\boldsymbol{r},\mu}\left(\mathrm{tr}\left[\phi^\dagger(\boldsymbol{r})\mathcal{U}_\mu(\boldsymbol{r})\phi(\boldsymbol{r} + \boldsymbol{e}_\mu)\right] + \text{c.c.}\right) \tag{18.8}$$

It is a simple excercise to show that in the naive continuum limit, $a_0 \to 0$, the Euclidean action becomes

$$S_{\text{matter}} = \int d^D x \, \frac{1}{2g_{\text{matter}}^2} \mathrm{tr}\left[\left(D_\mu\phi\right)^\dagger D_\mu\phi\right] \tag{18.9}$$

where D_μ is the covariant derivative, and $g_{\text{matter}}^2 \equiv (a_0^{D-2}/\beta)$.

The action for the gauge fields must also be locally gauge invariant. Hence, it must be written in terms of *Wilson loops*. Let $(\boldsymbol{r}; \mu, \nu)$ define an elementary oriented plaquette of the hypercubic lattice with vertex at the lattice site $\boldsymbol{r}$, and a face (the plaquette) with edges along the μ and ν directions (see Fig.18.2). The Wilson loop for an elementary plaquette with boundary γ is

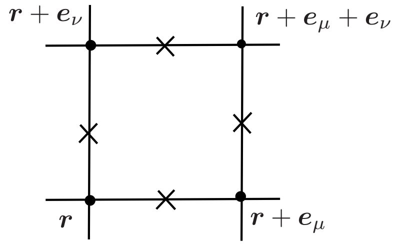

Figure 18.2 An elementary plaquette: the matter fields live on the sites (bold dots) and the gauge fields on the links (crosses).

$$W_\gamma = \operatorname{tr} \prod_{(\boldsymbol{x},\mu)\in\gamma_{\boldsymbol{r};\mu\nu}} \mathcal{U}_\mu(\boldsymbol{x}) \tag{18.10}$$

where the product runs on the links of the oriented plaquette $(\boldsymbol{r};\mu,\nu)$.

The gauge-invariant Wilson action for the gauge fields is

$$-S_{\text{gauge}} = \frac{1}{2g}\sum_{\boldsymbol{r};\mu\nu} \operatorname{tr}\Big(\mathcal{U}_\mu(\boldsymbol{r})\mathcal{U}_\nu(\boldsymbol{r}+\mathrm{e}_\mu)\mathcal{U}_\mu^{-1}(\boldsymbol{r}+\mathrm{e}_\nu)\mathcal{U}_\nu^{-1}(\boldsymbol{r}) + \text{c.c.}\Big) \tag{18.11}$$

where g is the coupling constant. It is elementary to show that in the naive continuum limit, $a_0 \to 0$ it becomes the Yang-Mills action,

$$S_{\text{gauge}} = \frac{1}{4\tilde{g}}\int d^D x \operatorname{tr} F_{\mu\nu}^2 \tag{18.12}$$

where we have set $\tilde{g} = a_0^{D-2} g$.

An important physical observable that we will analyze is the expectation value of a Wilson loop operator on a large closed loop Γ of the lattice,

$$\Big\langle W_\Gamma \Big\rangle = \Big\langle \operatorname{tr} \prod_{(\boldsymbol{x},\mu)\in\Gamma} \mathcal{U}_\mu(\boldsymbol{x}) \Big\rangle \tag{18.13}$$

which in the naive continuum limit, $a_0 \to 0$, is

$$\Big\langle W_\Gamma \rangle \sim \langle \operatorname{tr}\Big[P\exp\Big(i \oint_\Gamma dx_\mu A^\mu(x)\Big)\Big]\Big\rangle \tag{18.14}$$

It is implicit in this notation that, in a gauge theory with gauge group G, the Wilson loop carries the quantum numbers of a representation of the gauge group. Since the Wilson loop represents the coupling to a static, infinitely heavy, matter field, these are the quantum numbers carried by the matter field. We will see that in theories in which matter and gauge gauge fields are both dynamical, the behavior of this observable depends of the representation, i.e. depends on the *charge* of the matter fields. A general compact gauge group G has many such non-trivial representations.

In what follows, we will assume that, unless we state the contrary, that the Wilson loop carries the *fundamental* charge of the gauge group, i.e. the quantum numbers of the lowest non-trivial irreducible representation of the gauge group.

18.4 Gauge Fields

The partition function of a gauge theory on a gauge group G is

$$Z = \int \prod_{\boldsymbol{r},\mu} d\mathcal{U}_\mu(\boldsymbol{r}) \exp\left[\frac{1}{2g}\sum_{\boldsymbol{r};\mu\nu} \mathrm{tr}\left(\mathcal{U}_\mu(\boldsymbol{r})\mathcal{U}_\nu(\boldsymbol{r}+\boldsymbol{e}_\mu)\mathcal{U}_\mu^{-1}(\boldsymbol{r}+\boldsymbol{e}_\nu)\mathcal{U}_\nu^{-1}(\boldsymbol{r}) + \text{c.c.}\right)\right] \tag{18.15}$$

where we have used the Wilson action. Here dU denotes the invariant (Haar) measure of the group G. As in the continuum case, this partition function requires gauge fixing to have an integration measure that sums over gauge classes of configurations.

Examples of gauge theories are,

$\mathbb{Z}_2$: This is the Ising gauge theory, whose gauge group is $\mathbb{Z}_2$. We will use the notation $\mathcal{U}_\mu \equiv \tau_\mu$, with $\tau_\mu = \pm 1$. The partition function of the Ising gauge theory is

$$Z_{\text{Z2GT}} = \sum_{\tau_\mu=\pm 1} \exp\left(\frac{1}{g}\sum_{\boldsymbol{r},\mu\nu} \tau_\mu(\boldsymbol{r})\tau_\nu(\boldsymbol{r}+\boldsymbol{e}_\mu)\tau_\mu(\boldsymbol{r}+\boldsymbol{e}_\nu)\tau_\nu(\boldsymbol{r})\right) \tag{18.16}$$

where g is the coupling constant.

$U(1)$: This theory is known as compact electrodynamics. In this case $\mathcal{U}_\mu \in U(1)$ and can be written as $\mathcal{U}_\mu(\boldsymbol{r}) = \exp(iA_\mu(\boldsymbol{r}))$, where $A_\mu \in [0, 2\pi)$. The partition function is

$$Z_{U(1)} = \prod_{\boldsymbol{r};\mu,\nu} \int_0^{2\pi} \frac{dA_\mu(\boldsymbol{r})}{2\pi} \exp\left[\frac{1}{g}\sum_{\boldsymbol{r},\mu\nu} \cos(\Delta_\mu A_\nu(\boldsymbol{r}) - \Delta_\nu A_\mu(\boldsymbol{r}))\right] \tag{18.17}$$

where $\Delta\varphi(\boldsymbol{r}) \equiv \varphi(\boldsymbol{r}+\mathrm{e}_\mu) - \varphi(\boldsymbol{r})$ is the lattice right difference operator. Here too we write $K = 1/g$. The Maxwell theory is both the limit in which $a_0 \to 0$, *and* in which the fields A_μ are not compact, and hence take values on $\mathbb{R}$.

$SU(N)$: This is the lattice version of the Yang-Mills gauge theory with gauge group $SU(N)$. Here the lattice degrees of freedom ar group elements,

$\mathcal{U}_\mu \in SU(N)$. The action is given in Eq.(18.11), and the partition function is a sum over gauge field configurations with a Haar invariant measure (and suitably gauge-fixed).

We will not attempt to present a detailed version of what is known about these theories. Rather, we will discuss their *phase diagrams* and classify their possible ground states and the behavior of physical observables.

18.5 Hamiltonian Theory

18.5.1 Global Symmetries

In Chapter 14 we introduced the concept of the transfer matrix in Statistical Mechanics and showed that it is the analog of the Euclidean evolution operator. Using that approach we related the classical Ising model in D dimensions to a quantum Ising model in $d-1$ dimensions. The Hamiltonian of the quantum model has, in addition to the Ising interaction term, a transverse field term which plays the role of a kinetic energy. It is easy to see that this is a general result.

Let us summarize how this works. We define a hilbert space of states on a hypersurface at a fixed (discrete) imaginary time. The transfer matrix generates a new configuration (a new state) at the next time slice. In this way a configuration of the degrees of freedom in D dimensions is generated, and the partition function is a sum over such configurations. For instance, in the case of the Ising model, the Hilbert space is the tensor product of spin configurations at fixed time, and the resulting Hamiltonian is

$$H_{\text{Ising}} = -\sum_{\boldsymbol{r}} \sigma_1(\boldsymbol{r}) - \lambda \sum_{\boldsymbol{r},j} \sigma_3(\boldsymbol{r})\sigma_3(\boldsymbol{r}+\boldsymbol{e}_j) \tag{18.18}$$

where $j = 1, \ldots, D-1$, and σ_1 and σ_3 are Pauli matrices acting on the spin states. Here λ is the coupling constant. In this picture the low temperature phase of the classical theory is the large λ regime, and the high temperature phase of the classical theory is the small λ phase of the quantum theory.

Likewise, in the XY model, which has a global $U(1)$ symmetry, the degrees of freedom are phases, $\theta(\boldsymbol{r}) \in [0, 2\pi)$, and $L(\boldsymbol{r})$ is the conjugate momentum, such that $[\theta(\boldsymbol{r}), L(\boldsymbol{r}')] = i\delta_{\boldsymbol{r},\boldsymbol{r}'}$. The Hamiltonian now is

$$H_{XY} = \frac{1}{2}\sum_{\boldsymbol{r}} L^2(\boldsymbol{r}) - \frac{1}{g}\sum_{\boldsymbol{r},j} \cos\left(\Delta_j \theta(\boldsymbol{r})\right) \tag{18.19}$$

where, as before, $\Delta_j\theta(\boldsymbol{r}) = \theta(\boldsymbol{r}+\boldsymbol{e}_j) - \theta(\boldsymbol{r})$, and g is the coupling constant.

This construction generalizes to theories with non-abelian global symmetries, such as the $O(3)$ non-linear sigma model. In that case the degrees of freedom are labeled by configurations of vectors of unit length, $\boldsymbol{n}$, at each lattice site. The quantum lattice Hamiltonian is

$$H_{\text{NLSM}} = \frac{1}{2}\sum_{\boldsymbol{r}} \boldsymbol{J}^2(\boldsymbol{r}) - \frac{1}{g}\sum_{\boldsymbol{r},j} \boldsymbol{n}(\boldsymbol{r})\cdot\boldsymbol{n}(\boldsymbol{r}+\boldsymbol{e}_j) \tag{18.20}$$

where $J_i(\boldsymbol{r})$ are the generators of the group of rotations $O(3)$, which satisfy the commutation relations $[J_i(\boldsymbol{r}), J_j(\boldsymbol{r}')] = i\epsilon_{ijk}J_k(\boldsymbol{r})\delta_{\boldsymbol{r},\boldsymbol{r}'}$, and $[J_i(\boldsymbol{r}), n_j(\boldsymbol{r}')] = i\epsilon_{ijk}n_k(\boldsymbol{r})\delta_{\boldsymbol{r},\boldsymbol{r}'}$, and $\boldsymbol{J}^2$ is the quadratic Casimir operator.

In all three examples, in general dimension $D > D_c$ (where the critical dimension D_c depends on the whether the symmetry group is discrete or continuous), these theories have a symmetric phase and a phase with a spontaneously broken global symmetry, separated by a phase transition. In each phase there is a power series expansion with a radius of convergence smaller than the critical value of the coupling constant for the phase transition.

18.5.2 Local Symmetries: Gauge Theory

Something similar works for a gauge theory as well, with one important difference. In Chapter 9 we discussed the canonical quantization of gauge theories. There, we saw that the space of states has to be projected onto to the physical subspace of gauge-invariant states. This condition was implemented as a constraint that the physical states obey the Gauss law which plays the role of the generator of time-independent gauge transformations. The lattice theory has exactly the same structure.

We begin with the $\mathbb{Z}_2$ gauge theory. The analog of the temporal gauge is to set the $\mathbb{Z}_2$ gauge fields on the imaginary time direction to 1. Let $\sigma_j^1(\boldsymbol{r})$ and $\sigma_j^3(\boldsymbol{r})$ be two Pauli matrices defined on each link $(\boldsymbol{r}, j)$. The Hilbert space are the states of the $\mathbb{Z}_2$ gauge fields on each link of the lattice, e.g. the eigenstates of σ_3. The transfer matrix construction leads to the quantum Hamiltonian is now a sum of two terms, one on links (the "kinetic energy") and the other on plaquettes (the "potential energy"). It has the form

$$H_{\mathbb{Z}_2\text{GT}} = -\sum_{\boldsymbol{r},j}\sigma_j^1(\boldsymbol{r}) - \lambda\sum_{\boldsymbol{r};jk}\sigma_j^3(\boldsymbol{r})\sigma_k^3(\boldsymbol{r}+\boldsymbol{e}_j)\sigma_j^3(\boldsymbol{r}+\boldsymbol{e}_k)\sigma_k^3(\boldsymbol{r}) \tag{18.21}$$

Let us define at every site $\boldsymbol{r}$ the operators $Q(\boldsymbol{r})$

$$Q(\boldsymbol{r}) = \prod_j \left(\sigma_j^1(\boldsymbol{r})\sigma_{-j}^1(\boldsymbol{r})\right) \tag{18.22}$$

where, $\sigma^1_{-j}(\boldsymbol{r}) = \sigma^1_j(\boldsymbol{r}-\boldsymbol{e}_j)$. It is easy to see that $Q(\boldsymbol{r})$ flips the values of σ^3 on all the links that share the site $\boldsymbol{r}$ and, hence, is the generator of local $\mathbb{Z}_2$ gauge transformations. It is also apparent that it commutes with every term of the Hamiltonian, $[Q(\boldsymbol{r}), H] = 0$ and that they commute with each other, $[Q(\boldsymbol{r}), Q(\boldsymbol{r}')] = 0$. Hence, $Q(\boldsymbol{r})$ is the generator of time-independent $\mathbb{Z}_2$ gauge transformations, and plays the role of a (suitably generalized) Gauss law. Furthermore, since the operators $Q(\boldsymbol{r})$ are hermitian and obey $Q(\boldsymbol{r})^2 = 1$, their eigenvalues are just ± 1. Therefore, the quantum theory is defined on the physical space of gauge-invariant states $|\text{Phys}\rangle$, and obey

$$Q(\boldsymbol{r})|\text{Phys}\rangle = |\text{Phys}\rangle \tag{18.23}$$

that are invariant under local $\mathbb{Z}_2$ gauge transformations. States for which, at some value of $\boldsymbol{r}$, $Q(\boldsymbol{r}) = -1$ will be viewed as having $\mathbb{Z}_2$ sources located at $\boldsymbol{r}$. We will shortly see that, for any spacetime dimension $D > 2$, the $\mathbb{Z}_2$ gauge theory has two phases: a confining phase (for $\lambda < \lambda_c$), and a deconfined phase (for $\lambda > \lambda_c$).

Let us now move on to the case of a gauge theory with gauge group $U(1)$, known as compact electrodynamics. Here we will work in the temporal gauge, $A_0 = 0$. In this case, the degrees of freedom are vector potentials defined on the links of the lattice taking the values $A_j(\boldsymbol{r}) \in [0, 2\pi)$. The conjugate momenta are also defined on the links, are electric fields $E_j(\boldsymbol{r})$, which obey the commutation relations $[A_j(\boldsymbol{r}), E_k(\boldsymbol{r}')] = i\delta_{\boldsymbol{r},\boldsymbol{r}'}$. Thus, in this theory, electric fields behave as angular momenta and its eigenvalues are integer numbers. The Hamiltonian for the $U(1)$ gauge theory is

$$H_{U(1)\text{GT}} = \sum_{\boldsymbol{r},j} \frac{1}{2} E_j^2(\boldsymbol{r}) - \frac{1}{g} \sum_{\boldsymbol{r};jk} \cos\left(\Delta_j A_k(\boldsymbol{r}) - \Delta_k A_j(\boldsymbol{r})\right) \tag{18.24}$$

This theory is invariant under local gauge transformations. The local generators are written in terms of the lattice divergence operator as

$$Q(\boldsymbol{r}) \equiv \Delta_j E_j(\boldsymbol{r}) \equiv \sum_j \left(E_j(\boldsymbol{r}) - E_j(\boldsymbol{r}-\boldsymbol{e}_j)\right) \tag{18.25}$$

It is easy to see that the unitary transformations $\exp(i\sum_{\boldsymbol{r}} \theta(\boldsymbol{r})Q(\boldsymbol{r}))$ generate local gauge transformations $A_j(\boldsymbol{r}) \to A_j(\boldsymbol{r}) + \Delta_j\theta(\boldsymbol{r})$. Moreover, the operators $Q(\boldsymbol{r})$ commute with each other, $[Q(\boldsymbol{r}), Q(\boldsymbol{r}')] = 0$, and with the Hamiltonian. Hence, they are the generators of time-independent gauge transformations. We can now define the physical Hilbert space of states, $|\text{Phys}\rangle$, as the eigenstates of $Q(\boldsymbol{r})$ with zero eigenvalue,

$$Q(\boldsymbol{r})|\text{Phys}\rangle \equiv \Delta_j E_j(\boldsymbol{r})|\text{Phys}\rangle = 0 \tag{18.26}$$

which is the analog of the Gauss law constraint of Maxwell's electrodynamics. The main (and important!) difference is that in compact electrodynamics the operator $Q(\boldsymbol{r})$ has a spectrum of integer-valued eigenvalues. This implies that states that satisfy $Q(\boldsymbol{r})|\Psi\rangle = n(\boldsymbol{r})|\Psi\rangle$ have the interpretation of having sources that carry integer-valued electric charges. Therefore, in this theory, states with fractional charges are not allowed.

The non-abelian gauge theories are constructed in a similar fashion. here, for simplicity, we will consider the case of a theory with the $SU(2)$ gauge group. In this case the Hamiltonian is (using a conventional notation)

$$\begin{aligned} H_{SU(2)\mathrm{GT}} =& \frac{g}{2} \sum_{\boldsymbol{r},j} E_j^a(\boldsymbol{r}) E_j^a(\boldsymbol{r}) \\ & - \frac{1}{2g} \sum_{\boldsymbol{r};jk} \left[\mathrm{tr} \left(\mathcal{U}_j(\boldsymbol{r}) \mathcal{U}_k(\boldsymbol{r}+\boldsymbol{e}_j) \mathcal{U}_j^{-1}(\boldsymbol{r}+\boldsymbol{e}_k) \mathcal{U}^{-1}(\boldsymbol{r}) \right) + \mathrm{c.c.} \right] \end{aligned} \tag{18.27}$$

where, as before, the degrees of freedom are group elements, $\mathcal{U}_j(\boldsymbol{r}) \in G$, and $E_j^a(\boldsymbol{r})$ (with $a = 1, \ldots, D(G)$) are the generators of the Lie group G. In the case $G = SU(2)$, the group elements $\mathcal{U}_j(\boldsymbol{r})$ are 2×2 unitary matrices, the generators $E_j^a(\boldsymbol{r})$ are the angular momentum operators, and $E_j^a(\boldsymbol{r}) E_j^a(\boldsymbol{r})$ is the quadratic Casimir operatir. The generators $E_j^a(\boldsymbol{r})$ and the group elements $\mathcal{U}_j(\boldsymbol{r})$ satisfy the commutation relations

$$[E_j^a(\boldsymbol{r}), E_k^b(\boldsymbol{r}')] = i\epsilon^{abc} E_j^c(\boldsymbol{r}) \delta_{\mathrm{r},\boldsymbol{r}'} \delta_{jk}, \tag{18.28}$$

where ϵ^{abc} are the structure constants of $SU(2)$, and

$$[E_j^a(\boldsymbol{r}), \mathcal{U}_k(\boldsymbol{r}')] = \frac{1}{2} \sigma^a \mathcal{U}_j(\boldsymbol{r}) \delta_{jk} \delta_{\boldsymbol{r},\boldsymbol{r}'}, \tag{18.29}$$

where σ^a are three Pauli matrices (the elements of the algebra of the group $SU(2)$).

The generator of time-independent gauge transformations is, $Q^a(\boldsymbol{r})$ now takes values on the algebra of the group,

$$Q^a(\boldsymbol{r}) = \sum_{j=1}^{d} \left(E_j^a(\boldsymbol{r}) - E_j^a(\boldsymbol{r} - \boldsymbol{e}_j) \right) \tag{18.30}$$

and the Gauss law condition is

$$Q^a(\boldsymbol{r}) |\mathrm{Phys}\rangle = 0 \tag{18.31}$$

and defines the subspace of physical states.

Here too, states that are not annihilated by $Q^a(\boldsymbol{r})$ are viewed as not being in the vacuum sector but created by external probe sources. However, in a non-abelian theory, the Gauss-law generators do not commute with each other, and one can only specify the eigenvalues of the diagonal generators (i.e. in the Cartan sub-algebra) and the eigenvalue of the Casimir. For instance, in the $SU(2)$ gauge theory the sources carry the quantum numbers (j, m), where $j(j+1)$ is the eigenvalue of the Casimir and m is the eigenvalue of the diagonal generator of $SU(2)$. Hence, here too, the sources can only have a definite set of possible eigenvalues.

18.6 Elitzur's Theorem and the Observables of a Gauge Theory

There is a fundamental result known as elitzur's Theorem. It states that in a gauge theory with a compact gauge group, the only operators that may have a non-vanishing expectation values must be invariant under local gauge transformation. This theorem holds always, regardless of the phase in which the gauge theory is in.

This theorem stands in striking contrast with the case of theories with global symmetries. In that case we saw that if the global symmetry group is compact, the ground state may break this symmetry spontaneously. The direct implication of Elitzur's Theorem is that a local gauge symmetry *cannot be broken spontaneously*. This result may seem surprising since there are regimes of theories, such as the ones in which the Higgs mechanism is operative, in which it is often stated (incorrectly) that the gauge symmetry is spontaneously broken. We will clarify this apparent contradiction shortly.

To understand the meaning of Elitzur's Theorem, let us first revisit the assumptions behind the concept of spontaneous symmetry breaking. In that case we have a theory with a global compact symmetry group G. We assume that the theory is in a finite but large volume V with an extensive number of degrees of freedom. The prototype is an Ising model, which has two states per site, on a lattice with N^D sites. The action of the theory is invariant under global G transformations.

Then, one considers a theory in which a symmetry breaking field, with strength h, is added to the action. The theory now is no longer invariant under the global symmetry transformations G. For the sake of definiteness, let us think of an Ising model, which has a global $\mathbb{Z}_2$ symmetry. In this case, h represents an external uniform magnetic field. Spontaneous symmetry breaking here means that there is a non-vanishing local order parameter. Consider now a *finite* system ($V < \infty$) with a fixed value of the symmetry breaking field h. The presence of the symmetry breaking field causes the action to increase for configurations in which the majority of the degrees of freedom are in direction opposite with respect to the symmetry breaking field. Their contribution to the partition function is the suppressed.

If we now lower the temperature T (i.e. the coupling constant) below some critical value T_c, two things may happen, a) if h is reduced all the way to zero at fixed temperature and volume, then, in this case, the misaligned configurations will not be suppressed, and the average magnetization will vanish and the symmetry is not broken, b) we take the thermodynamic limit first ($V \to \infty$); this has the effect of imposing an infinite penalty on the misaligned configurations; if the symmetry breaking field is then turned off, the local magnetization, which is odd under the symmetry, will have a non-vanishing expectation value whose sign is determined by the now suppressed symmetry breaking field h. Therefore, the order of limits between the size of the system and the symmetry breaking field, matters. We saw this same result when we derived Goldstone's theorem in Chapter 12.

Another way to understand this result is in terms of perturbation theory: in the broken symmetry phase, the order in perturbation theory needed to mix two states that are related by the global symmetry is proportional to the size of the system. In that case, if the thermodynamic limit is taken first, the configurations related by the global symmetry will never mix. In other words, spontaneous breaking of a global symmetry is possible because the states that are related by the action of the global symmetry in the thermodynamic limit are infinitely far apart from each other. A consequence of this fact is that systems with a global symmetry are sensitive to the effects of symmetry breaking fields (even for infinitesimal ones) in the thermodynamic limit.

The situation is completely different in the case of a gauge theory. Consider, again for simplicity, the $\mathbb{Z}_2$ gauge theory in the Hamiltonian picture. Suppose we begin with some state $|\Psi\rangle$ that is not invariant under gauge transformations, i.e. such that for some sites $\{\boldsymbol{r}\}$, $Q(\boldsymbol{r})|\Psi\rangle = |\Phi\rangle \neq |\Psi\rangle$. Since the operator $Q(\boldsymbol{r})$ (i.e. the Gauss law) is the generator of time-independent local gauge transformations, the states $|\Psi\rangle$ and $|\Phi\rangle$ differ by a local gauge transformation. However, since the gauge group (in this case $\mathbb{Z}_2$) is compact, the orbit of locally inequivalent of states is finite. This implies that after acting a finite number of times on the images of $|\Psi\rangle$ one will find a state that will have a finite overlap with $|\Psi\rangle$. Hence, this process does not involve acting all the way to the boundaries and, in fact, holds for an infinitely large system. The argument given above holds also for continuous compact groups (such as $U(1)$, $SU(N)$, etc.) since in all cases the orbit of the gauge group has a finite measure.

Another, more formal, way to obtain this result is to consider, again for simplicity, a $\mathbb{Z}_2$ gauge theory on a hypercubic (Euclidean) space-time lattice in D dimensions. The following arguments are generalized with minor changes to theories with an arbitrary compact gauge group. Here we will consider a modified theory which has a term on the links of the lattice, of strength h, that breaks the local $\mathbb{Z}_2$ gauge invariance. The partition function of the modified $\mathbb{Z}_2$ gauge theory is (in a compact notation)

$$Z[K,h] = \mathrm{Tr}\exp\left(K\sum_{\text{plaquettes}}\tau_\mu\tau_\nu\tau_\mu\tau_\nu + h\sum_{\text{links}}\tau_\mu\right) \tag{18.32}$$

where $K = 1/g$ (g being the coupling constant), and the trace is the sum over all the configurations of the $\mathbb{Z}_2$ gauge fields, here denoted by $\tau_\mu = \pm 1$ on each link (of direction μ) of the lattice. The integration measure (i.e. the sum over the gauge field configurations) and the action of this theory, except for the presence of the symmetry breaking field h, are invariant under local $\mathbb{Z}_2$ gauge transformations.

Let us consider the computation of the expectation value of a gauge non-invariant operator, such as $\tau_\mu(\boldsymbol{r})$ located on a *single link* $(\boldsymbol{r},\mu)$,

$$\langle\tau_\mu(\boldsymbol{r})\rangle_{K,h} = \frac{1}{Z[K,h]}\mathrm{Tr}\left[\tau_\mu(\boldsymbol{r})\exp\left(K\sum_{\text{plaquettes}}\tau_\mu\tau_\nu\tau_\mu\tau_\nu + h\sum_{\text{links}}\tau_\mu\right)\right] \tag{18.33}$$

Let us consider a local gauge transformation $G(\boldsymbol{r})$ of the gauge fields $\tau_\mu(\boldsymbol{r})$ on the links emanating from a site $\boldsymbol{r}$ (shown in Fig.18.3)

$$\tau_\mu(\boldsymbol{r}) \mapsto \tau'_\mu(\boldsymbol{r}) = -\tau_\mu(\boldsymbol{r}), \quad \forall \text{ links } (\boldsymbol{r},\mu) \text{ attached to site } \boldsymbol{r} \tag{18.34}$$

All other gauge fields are unaffected by the local transformation $G(\boldsymbol{r})$. We

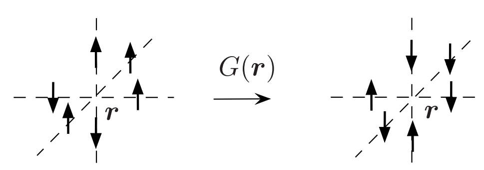

Figure 18.3 A $\mathbb{Z}_2$ gauge transformation at site $\boldsymbol{r}$, denoted bt $G(\boldsymbol{r})$: the gauge field variables (denoted by arrows) on all the links attached to the site $\boldsymbol{r}$ are flipped.

next perform a change of variables of the gauge fields, $\tau_\mu(\boldsymbol{r}) \mapsto \tau'_\mu(\boldsymbol{r})$, that coincides with the gauge transformation $G(\boldsymbol{r}$ of Eq.(18.34) on the links attached to the site $\boldsymbol{r}$, and unchanged otherwise. making the change in the expectation value of Eq.(18.33), we find

$$\begin{aligned}\langle \tau_\mu(\boldsymbol{r}) \rangle =& -\frac{1}{Z[K,h]}\mathrm{Tr}\left[\tau'_\mu(\boldsymbol{r}) \exp\left(K \sum_{\text{plaquettes}} \tau'_\mu \tau'_\nu \tau'_\mu \tau'_\nu + h \sum_{\text{links}} \tau'_\mu \right)\right] \\ =& \Big\langle -\tau_\mu(\boldsymbol{r}) \exp\Big(-2h \sum_{\{(\boldsymbol{r},\nu)\}} \tau_\nu(\boldsymbol{r}) \Big) \Big\rangle_{K,h}\end{aligned} \tag{18.35}$$

where $\{(\boldsymbol{r},\nu)\}$ is the set of links attached to the site $\boldsymbol{r}$, and the expectation value is taken in the theory with couplings K and h.

Next, one observes that the following bound on a local change holds,

$$\Big| \Big\langle \tau(\boldsymbol{r},\mu) \Big\rangle_{K,h} - \Big\langle -\tau_\mu(\boldsymbol{r},\mu) \Big\rangle_{K,h} \Big| = \Big| \Big\langle -\tau_\mu(\boldsymbol{r}) \Big[\exp\Big(-2h \sum_{\{(\boldsymbol{r},\nu)\}} \tau_\nu(\boldsymbol{r}) \Big) - 1 \Big] \Big\rangle_{K,h} \Big| \tag{18.36}$$

where D is the dimension. The expression on the left hand side of this identity, Eq.(18.36), has the upper bound

$$\Big| \Big\langle -\tau_\mu(\boldsymbol{r}) \Big[\exp\Big(-2h \sum_{\{(\boldsymbol{r},\nu)\}} \tau_\nu(\boldsymbol{r}) \Big) - 1 \Big] \Big\rangle_{K,h} \Big| \leq (e^{4Dh} - 1) \Big| \langle \tau_\mu(\boldsymbol{r},\mu) \rangle_{K,h} \Big| \tag{18.37}$$

Hence, the left hand side of Eq.(18.36) has the same upper bound. However, since as $h \to 0$ this upper bound approaches zero. Thus

$$\lim_{h\to 0} \langle \tau_\mu(\boldsymbol{r},\mu) \rangle_{K,h} = \lim_{h\to 0} \langle -\tau_\mu(\boldsymbol{r},\mu) \rangle_{K,h} \tag{18.38}$$

this implies that, in the limit $h \to 0$, the expectation value of this gauge non-invariant operator (and, in fact, of *any* gauge non-invariant operator) must vanish,

$$\langle \tau_\mu(\boldsymbol{r},\mu) \rangle_K = 0 \tag{18.39}$$

We note that this proof has two key ingredients: a) that the transformation is local, and b) that the change in the action due to the gauge transformation is finite. Note that, in contrast, in the case of a global symmetry the change in the action diverges in the thermodynamic limit.

Therefore, we conclude that only locally gauge invariant operators can have a non-vanishing expectation value. A direct consequence of this result is that *gauge symmetries cannot be spontaneously broken.* As stated above, this theorem holds for all theories with a local gauge invariance with a compact gauge group.

18.7 Phases of Gauge Theories

We will now discuss the phases that can exist in pure gauge theories. For simplicity we will ficus on the $\mathbb{Z}_2$ gauge theory and on compact electrodynamics. However we will comment what results are generic. We will consider both the Euclidean and the Hamiltonian versions whenever necessary.

18.7.1 *Weak Coupling: Deconfinement*

In the weak coupling limit, $g \to 0$, perturbation theory works. In this limit, the Gibbs weight for each plaquette of the D-dimensional hypercubic lattice (using a synthetic notation)

$$\exp\left[\frac{1}{2g}\mathrm{tr}\left(\mathcal{U}_\mu\mathcal{U}_\nu\mathcal{U}_\mu^{-1}\mathcal{U}_\nu^{-1}\right) + \text{c.c.}\right] \tag{18.40}$$

is dominated by flat gauge field configurations (i.e. such that $F_{\mu\nu} = 0$). Hence, in this extreme limit we can set $\mathcal{U}_\mu = I$ (up to gauge transformations). This is the phase we have studied so far. For simplicity we will consider the $\mathbb{Z}_2$ gauge theory whose partition function is

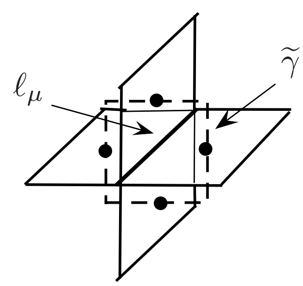

Figure 18.4 The leading term of the weak coupling expansion to the partition function of the $\mathbb{Z}_2$ gauge theory in $D = 3$ dimensions is the flipped link ℓ_μ. The four plaquettes (labelled by dots) that share this link are flipped. A flipped link can then be regarded as a closed *magnetic* loop on the dual lattice (shown as a closed broken curve). In $D = 4$ dimensions the dual magnetic loop is a surface and, more generally a $D - 2$-dimensional hypersurface.

$$Z[\mathbb{Z}_2] = \mathrm{Tr}\exp\Big(\sum_{\text{plaquettes}}\frac{1}{g}\tau_\mu\tau_\nu\tau_\mu\tau_\nu\Big) \tag{18.41}$$

where trace means the sum over all configurations of gauge fields $\tau_\mu = \pm 1$ (modulo gauge fixing).

In the ultra-weak coupling regime, $g \to 0$, we can start from the configuration in which $\tau_\mu = 1$ in all links (up to gauge transformations, and neglecting for now interesting topological issues arising on closed manifolds). In this configuration, the product of the four τ_μ gauge fields on each plaquette is equal to $+1$. A configuration with one flipped link variable to the value -1, flips the values of all plaquettes that share this link from $+1$ to -1. This increases the action by an amount of $2/g$ times the number of flipped plaquettes N_p (where $N_p = 4$ in $D = 3$ dimensions). In $D = 3$ dimensions the flipped plaquettes can be regarded as being threaded by a closed loop on the dual lattice, that we will call a magnetic loop, as shown in Fig.18.4. In $D > 3$ dimensions, the magnetic loop is a surface in $D = 4$ dimensions, and, in general, a $D - 2$ hypersurface in D dimensions. Thus, this configuration has a Gibbs weight of $\exp(-\frac{2}{g}N_p)$ times the number of places where we can flip a link. In the $\mathbb{Z}_2$ gauge theory, this expansion turns out to have a finite radius of convergence for $D > 2$.

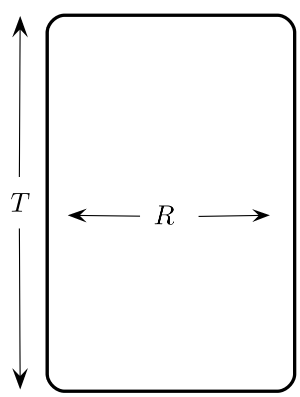

Figure 18.5 A Wilson loop Γ in Euclidean space-time. Here T is the imaginary time span and R is the spacial size of the loop.

For a $U(1)$ theory, the weak coupling regime is described by Maxwell's electrodynamics, whose excitations are photons. In fact, in section 9.7 we computed the expectation value of the Wilson loop operator in Euclidean space-time. There we found that the expectation value of a Wilson loop on a closed contour Γ, such the one shown in Fig.18.5,

$$\langle W_\Gamma \rangle = \exp\left(-TV(R)\right) \tag{18.42}$$

By explicit computation we showed that the effective potential $V(R)$ is just the Coulomb interaction

$$V_{\text{Maxwell}}(R) = \frac{e^2}{R} \tag{18.43}$$

where we have identified the coupling constant with the electric charge, $g = e$. Notice that, in this case, we could have chosen the charge of the Wilson loop to be ne, with $n \in \mathbb{Z}$, which would have only changed the result my a factor of n^2.

We will call this behavior the Coulomb Phase. We should note that on a lattice the Wilson loops are not smooth everywhere and have cusps, which yield singular contributions proportional to $\ln R$. We will ignore these singularities in what follows.

Notice that in $D = 4$ dimensions, the charge e is dimensionless, and the ratio T/R is invariant under dilatations of spacetime, $R \to \lambda R$ and $T \to \lambda T$. From this perspective, we should expect that at a fixed point if the renormalization group (i.e. in a scale-invariant theory) that the expectation value of the Wilson loop operator should be a universal function $F(x)$ of the scale-invariant ratio $x = R/T$. Hence, in *any dimension*, at a fixed point of a gauge theory we expect that the effective potential should obey a $1/R$ law.

On the other hand, the case of Yang-Mills theory in $D = 4$ dimensions, this phase, which would have massless quarks and gluons, is asymptotically free and IR unstable, e.g. the renormalization group beta function is $\beta(g) = -Ag^2 + O(g^3)$ (where A is a positive constant). In this case, the coupling constant flows to strong coupling in the IR, and weak coupling perturbation theory breaks down beyond a finite length scale ξ. We will see that this is the confinement scale. The expectation value of the Wilson loop exhibits a crossover from loops smaller than the confinement scale, where they behave much in the same way as in perturbative Yang-Mills theory (with a slowly scale dependent coupling constant), to an area law for large loops (as we will see below). State-of-the-art numerical Monte Carlo simulations in $D = 4$ dimensions have established that the area law behavior holds from the strong coupling regime down to a sufficiently weak coupling regime where perturbative renormalization group calculations are still reliable. These numerical results are strong evidence that confinement holds in the continuum field theory.

Gauge theories with a discrete gauge group have a different behavior. Indeed, in $D > 2$ dimensions, the $\mathbb{Z}_2$ gauge theory has a convergent weak coupling expansion for $g < g_c$. This expansion is reminiscent to the low temperature expansion in a classical spin system. Indeed, in this theory too,

$g = 0$ implies that only flat configurations of the gauge field will contribute. This means that the the product of the τ_μ gauge variables one each plaquette should be equal to 1. Up to gauge transformations this is equivalent to set $\tau_\mu = 1$ on all the links of the lattice.

Let us compute the expectation value of the Wilson loop operator in the weak coupling phase of the $\mathbb{Z}_2$ gauge theory. This theory has only one non-trivial representation and, hence, there is only one $\mathbb{Z}_2$ charge. Clearly, in this limit, the expectation value of the Wilson loop is trivially $\langle W_\gamma \rangle = 1$. The first excitation above this state consists on flipping the value of τ_μ from $+1$ to -1 on just one link. However, this flips the value of the product of the $\mathbb{Z}_2$ gauge fields on all the plaquettes that share this link. The number of such plaquettes is $2(D-1)$. Therefore, the cost in the action of flipping just one τ variable is $4(D-1)$. For only one flip, for a lattice of N sites and a loop γ of perimeter $L = 2(T+R)$, the expectation value of the Wilson loop is

$$\langle W_\Gamma \rangle = \frac{1 + (N-2L)\exp(-4(D-1)/g) + \dots}{1 + N\exp(-4(D-1)/g) + \dots} \tag{18.44}$$

Let us consider the case of n flipped links. Provided g is small enough that we can, to a first approximation, we can ignore configurations with adjacent flipped links, and we can treat the flipped links as being dilute. In this approximation, the nth order correction to the numerator of Eq.(18.44) is

$$\frac{1}{n!}(N-2L)^n \exp(-4n(D-1)/g) \tag{18.45}$$

In the same approximation, the denominator (i.e. the partition function Z) has a contribution

$$\frac{1}{n!}N^n \exp(-4n(D-1)/g) \tag{18.46}$$

Therefore, in the dilute limit, the partition function is

$$\begin{aligned} Z \simeq& 1 + N\exp(-4(D-2)/g) + \dots + \frac{1}{n!}N^n \exp(-4n(D-1)/g) + \dots \\ =& \exp(N\exp(-4(D-1)/g)) \end{aligned} \tag{18.47}$$

Within the same approximation, the numerator of Eq.(18.44) now becomes

$$\begin{aligned} &1 + (N-2L)\exp(-4(D-1)/g) + \dots + \frac{1}{n!}(N-2L)^n \exp(-4n(D-1)/g) + \dots \\ &= \exp\left((N-2L)\exp(-4(D-1)/g)\right) \end{aligned} \tag{18.48}$$

Therefore, to this order of approximation, we find

$$\langle W_\Gamma \rangle = \exp(-2\exp(-4(D-1)/g)L) \tag{18.49}$$

In other words, in the limit of g small enough, the Wilson loop obeys a *perimeter law* of the form

$$\langle W_\Gamma \rangle = \exp(-f(g)L) \tag{18.50}$$

To the order of approximation that we have used, $f(g) = 2\exp(-4(D-1)/g) + \ldots$, up to exponentially small corrections in the spatial extent R of the loop. In this regime, the effective potential becomes

$$V(R) = \lim_{R/T \to 0} 4\exp(-2(D-1)/g)(1 + R/T) = 4\exp(-2(D-1)/g) + \ldots \tag{18.51}$$

which, up to exponentially small additive corrections, is a constant. Thus, to this order of approximation, the effective potential $V(R)$ is simply twice the self-energy cost of each source. Since there is no dependence on R, at this order their effective interaction potential $V_{\text{int}}(R) \approx 0$. One can prove that this is the actually the leading behavior (for a very large loop) of an expansion that has a finite radius of convergence. The leading non-vanishing contribution to the interaction potential decays exponentially fast with R, i.e. $\exp(-R/\xi)$, where ξ is the range of the effective interaction. These results also hold for other gauge theories with a discrete gauge group, e.g. $\mathbb{Z}_n$, provided the spacetime dimension is $D > 2$.

18.7.2 Strong Coupling and Confinement: Hamiltonian Picture

$\mathbb{Z}_2$ Gauge Theory

We begin the discussion of confinement in the strong coupling regime from the perspective of the Hamiltonian picture. We will consider two theories: the $\mathbb{Z}_2$ (Ising) gauge theory and the $U(1)$ (compact electrodynamics) gauge theory.

We consider first the Hamiltonian $\mathbb{Z}_2$ gauge theory in d space dimensions. We will write the Hamiltonian of the theory, Eq.(18.21), as the sum of two terms $H_{\mathbb{Z}_2} = H_0 + \lambda H_1$, where $\lambda = 1/g$ is the coupling constant, and where we set

$$H_0 = -\sum_{\boldsymbol{r},j} \sigma_j^1(\boldsymbol{r}), \quad H_1 = -\sum_{\boldsymbol{r};jk} \sigma_j^3(\boldsymbol{r})\sigma_k^3(\boldsymbol{r}+\boldsymbol{e}_j)\sigma_j^3(\boldsymbol{r}+\boldsymbol{e}_k)\sigma_k^3(\boldsymbol{r}) \tag{18.52}$$

The physical Hilbert space are the gauge-invariant states that satisfy

$$Q(\boldsymbol{r})|\text{Phys}\rangle = \prod_j \left(\sigma_j^1(\boldsymbol{r})\sigma_{-j}^1(\boldsymbol{r})\right)|\text{Phys}\rangle = |\text{Phys}\rangle \tag{18.53}$$

In the strong coupling limit, $g \gg 1$ (i.e. $\lambda \ll 1$), we can regard H_0 as the unperturbed Hamiltonian and H_1 as the perturbation. In this limit, the Hamiltonian H_0 is diagonal in the basis of the $\{\sigma_j^1(\mathrm{r})\}$ operators on the links. In this basis, the generators of local gauge transformations $\{Q(\boldsymbol{r})\}$ are also diagonal. Since acting on gauge-invariant states $Q(\boldsymbol{r}) = +1$, for all sites $\boldsymbol{r}$, in this basis gauge-invariant states must have, at most, an even number of links where $\sigma_j^1 = -1$ at every site of the lattice.

The unperturbed ground state, $|\mathrm{gnd}\rangle_0$ then must be the state with $\sigma_j^1 = +1$ on all links of the lattice. The energy of the unperturbed ground state is $E_{\mathrm{gnd}_0} = -Nd$. It is easy to see that, since the plaquette operators in H_1 flip the values of σ_j^1 around a plaquette from $+1$ to -1, the excited states are closed loops γ of the lattice where $\sigma_j^1 = -1$. In particular, for λ small enough (g large enough) only small "flipped" loops are present in the ground state.

We will now consider a different sector of the Hilbert space defined by the condition that the operators $\{Q(\boldsymbol{r})\}$ are equal to $+1$ everywhere except at two sites that we will denote by $\boldsymbol{R}$ and $\boldsymbol{R}'$, where $Q(\boldsymbol{R}) = Q(\boldsymbol{R}') = -1$. This choice is consistent with the requirement that the global $\mathbb{Z}_2$ symmetry must be unbroken, which requires that there must be an even number of $\mathbb{Z}_2$ sources. The eigenvalue of the operator $Q(\boldsymbol{r})$ is the $\mathbb{Z}_2$ 'electric' charge of the state.

For simplicity, we will choose $\boldsymbol{R}$ and $\boldsymbol{R}'$ to be on the same row of the lattice (see Fig.18.6a). We will say that, in this sector of the Hilbert space consists of states that have two $\mathbb{Z}_2$ "charges", $Q = -1$, one at $\boldsymbol{R}$, and the other at $\boldsymbol{R}'$. The ground state in this sector must have the smallest number of links with $\sigma_j^1 = -1$, and must obey that $Q = -1$ only at $\boldsymbol{R}$ and at $\boldsymbol{R}'$, and has $Q = +1$ on all other sites. It is obvious that the ground state in this sector, that we will denote by $|\boldsymbol{R}, \boldsymbol{R}'\rangle$, has $\sigma_j^1 = -1$ on the links of the shortest path γ on the lattice from $\boldsymbol{R}$ to $\boldsymbol{R}'$. In the case shown in Fig.18.6a, the shortest path γ is just the straight line (shown as the dark links in the figure).

The operator that creates this state is a *Wilson arc* of the form

$$W_\gamma[\boldsymbol{R}, \boldsymbol{R}'] = \prod_{(\boldsymbol{r},j)\in\gamma} \sigma_j^3(\boldsymbol{r}) \tag{18.54}$$

where γ is an open path on the links of the lattice with endpoints at $\boldsymbol{R}$ and $\boldsymbol{R}'$. It is easy to see that the operator $Q(\boldsymbol{r})$ commutes with the Wilson arc for all $\boldsymbol{r} \neq \boldsymbol{R}, \boldsymbol{R}'$, but they anticommute if $\boldsymbol{r} = \boldsymbol{R}, \boldsymbol{R}'$. Thus, this operator is not invariant under local gauge $\mathbb{Z}_2$ transformations, and is not allowed in the vacuum sector of the theory. However it is allowed in the sector where $Q(\boldsymbol{r}) = -1$ at $\boldsymbol{r} = \boldsymbol{R}, \boldsymbol{R}'$, i.e. in a sector with two static sources.

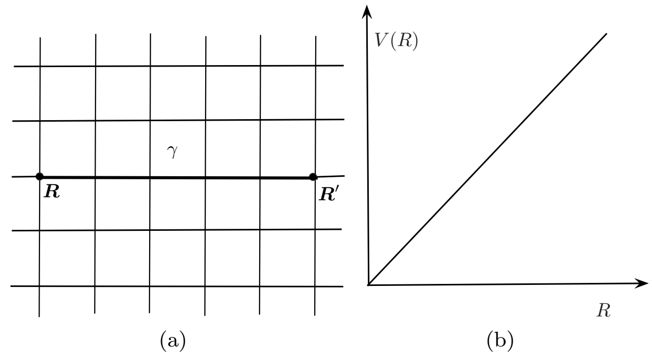

Figure 18.6 a) A string state in the strong coupling regime of the $\mathbb{Z}_2$ gauge theory: $\sigma_j^1 = -1$ on the dark links stretching from $\boldsymbol{R}$ to $\boldsymbol{R}'$ along the path γ. In this state $\sigma_j^1 = +1$ on all other links. b) Confining potential of two $\mathbb{Z}_2$ charges separated a distance $|\boldsymbol{R} - \boldsymbol{R}'|$ in the strong coupling regime.

The energy of this state is

$$E[\boldsymbol{R}, \boldsymbol{R}'] = E_{\text{gnd}_0} + 2|\boldsymbol{R} - \boldsymbol{R}'| \tag{18.55}$$

In other terms, the energy cost of introducing two $\mathbb{Z}_2$ sources at $\boldsymbol{R}$ and $\boldsymbol{R}'$ is the effective potential $V(|\boldsymbol{R} - \boldsymbol{R}'|)$

$$V(|\boldsymbol{R} - \boldsymbol{R}'|) = \sigma(g)|\boldsymbol{R} - \boldsymbol{R}'| \tag{18.56}$$

where

$$\sigma(g\lambda) = 2 - O(\lambda^2) \tag{18.57}$$

is the *string tension*. Hence, the energy needed to separate two $\mathbb{Z}_2$ charges grows *linearly* with their separation $|\boldsymbol{R} - \boldsymbol{R}'|$, and we say that the charges are *confined*, see Fig.18.6b.In other words, the two charges experience a force much like that of a string.

It is easy to see that the expansion in powers of $\lambda = 1/g$ has a finite radius of convergence. These corrections are virtual states on longer strings, and lead to a progressive reduction of the string tension σ as λ increases (g decreases). In other words, the string becomes progressively "fatter", with a transversal thickness ξ, such that $\xi \to 0$ as $\lambda \to 0$, and its energy per unit length progressively decreases. In this picture, the phase transition to the deconfined phase occurs when the string tension $\sigma \to 0$. It turns out (as we will see) that this is a continuous transition, i.e. $\sigma(\lambda) \propto |\lambda - \lambda_c|^\rho$ (here ρ is a critical exponent), in $D = 2 + 1$ space-time dimensions, but is discontinuous in higher dimensions.

On the other hand, in the deconfined phases, $g \gg 1$, the action of the the Wilson arc operator is quite simple. In this limit the plaquette operator, which is a product of the σ^3 link operators around each plaquette, plays the role of the unperturbed Hamiltonian, and the link operator σ^1 is the perturbation. In the extreme limit $g \to 0$, the natural basis (up to gauge fixing) for the Hilbert space are the eigenstates of σ^3 on the links. In this limit the ground state has $\sigma^3 = +1$ on each link (up to local gauge transformations). This state is not gauge invariant. One can fix the gauge, and then choose this state as the ground state. In this basis, the Wilson arc operator is diagonal and acts as a c-number.

Alternatively, one can work in the basis in which the $\mathbb{Z}_2$ charge operator is diagonal. This is the σ^1 basis, the one we used in the string coupling limit. We saw that in the strong coupling regime, in this basis the vacuum can be viewed as a linear superposition of states each of the form of a collection of closed strings (loops) on the lattice. In the confining phase the loops have finite size, typically of the order of the confinement scale. The phase transition to the deconfined (weak coupling) phase can be viewed as a proliferation of the loops, which become arbitrarily large. In the extreme deconfined phase, $g \to 0$, the vacuum state is the equal-amplitude linear superposition of configurations of all loops, regardless of their size. We will see shortly that this state embodies to topological nature of the deconfined phase.

In addition to these "electric excitations", this theory also has "magnetic excitations". In 2+1 dimensions the operator that creates a magnetic excitation flips the value of the local plaquette operator from +1 to -1. In $d = 2$ space dimensions this operator is similar to a magnetic monopole, and is defined as

$$M[\widetilde{\boldsymbol{R}}] = \prod_{(\boldsymbol{r},j)\in\tilde{\gamma}(\widetilde{\boldsymbol{R}})} \sigma_j^1(\boldsymbol{r}) \tag{18.58}$$

Hence, this operator is the product of σ^1 link operators a seam of links of the lattice pierced by an open curve $\tilde{\gamma}$ of teh dual of the square lattice with end-point at the dual site $\widetilde{\boldsymbol{R}}$, as shown in Fig.18.7. From its definition, it is clear that, in the deconfined phase, this gauge-invariant operator creates a pair of $\mathbb{Z}_2$ monopoles at the opposite ends of its "Dirac string" $\tilde{\gamma}$. These monopole states have finite energy and are deconfined. Thus, in its deconfined phase, the theory has electric and magnetic charges that have a finite energy gap. Nevertheless, both excitations are created by non-local operators. On the other hand, in the confined phase, $g \gg 1$, this operator acts as a c-number.

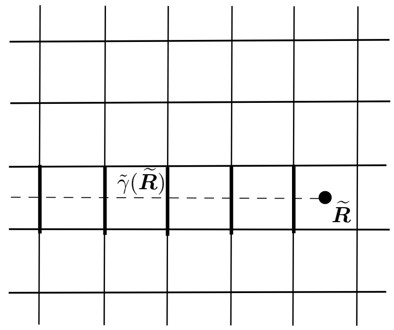

Figure 18.7 A "magnetic excitation" (or $\mathbb{Z}_2$ monopole) is created by a non-local operator that flips the value of the plaquette operator at the plaquette labelled by the dark dot. This operator is defined on a path $\tilde{\gamma}$ of the dual lattice.

Hence, it has a finite expectation value in the confined phase. In this sense, the confined phase of the $\mathbb{Z}_2$ gauge theory is a magnetic condensate.

With some important modifications, an analog of these operators exists in higher dimensions. While the Wilson arc can be defined in all dimensions, the analog of the magnetic operator changes with dimension. In $D = 3 + 1$ dimensions the magnetic operator is defined on a closed loop $\tilde{\gamma}$ of the dual lattice, and is known as a 't Hooft magnetic loop. The closed loop $\tilde{\gamma}$ on the dual (cubic) lattice is a closed curve on the dual lattice piercing a closed tube of plaquettes of the cubic lattice. The closed curve $\tilde{\gamma}$ is the boundary of an open surface $\tilde{\Sigma}$, with the topology of a disk, with the topology of a disk.

The 't Hooft magnetic loop operator, which we will denote by $\widetilde{W}[\tilde{\gamma}]$, is a product of σ^1 link operators on all the links of the lattice that pierce the plaquettes of the surface $\tilde{\Sigma}$ on the dual lattice. It was originally introduced by 't Hooft as a criterion for confinement in non-abelian gauge theory, where it acts on the center $\mathbb{Z}_N$ of the group $SU(N)$. In two space dimensions the 't Hooft operator (i.e. the monopole operator $M[\widetilde{\boldsymbol{R}}]$ defined above), is labeled by a point, in three space dimensions by a closed loop, in four space dimensions by a closed surface, which is the boundary of a three-volume, and so on. Returning to the case of $3 + 1$ dimensions, it is straightforward to see the the 't Hooft loop has an *area* law in the *deconfined* phase of the $\mathbb{Z}_2$ gauge theory, and a *perimeter* law in the *confined* phase. In other words, the 't Hooft and the Wilson loop operators exhibit opposite behaviors.

U(1) Gauge Theory

It is straightforward to see that the same behavior is seen on all other gauge theories with a compact gauge group G. To clarify how this happens we will consider briefly the case of a $U(1)$ gauge theory ("compact QED") also in the Hamiltonian picture. The Hamiltonian of the $U(1)$ gauge theory was given in Eq.(18.24). It can also be split into a sum of two terms, an unperturbed Hamiltonian H_0

$$H_0 = \sum_{\boldsymbol{r},j} \frac{1}{2} E_j^2(\boldsymbol{r}) \tag{18.59}$$

and a perturbation H_1,

$$H_1 = -\frac{1}{g} \sum_{\boldsymbol{r};jk} \cos\left(\Delta_j A_k(\boldsymbol{r}) - \Delta_k A_j(\boldsymbol{r})\right) \tag{18.60}$$

The physical Hilbert space are the states that satisfy Gauss Law, which now reads

$$Q(\boldsymbol{r})|\text{Phys}\rangle = \Delta_j E_j(\boldsymbol{r})|\text{Phys}\rangle = 0 \tag{18.61}$$

In the strong coupling regime, $g \gg 1$, H_1 is parametrically small and we can construct the states in a perturbative expansion in powers in $1/g$. In this regime, it is natural to work in the basis in which the electric field operators on the links, $E_j(\boldsymbol{r})$, are diagonal. Since the gauge fields take values in the compact group $U(1)$, i.e. $A_j(\boldsymbol{r}) \in [0, 2\pi)$, the eigenstates of the electric fields take values on the integers, $\ell_j(\boldsymbol{r}) \in \mathbb{Z}$, which label the representations of $U(1)$. In this description, the physical Hilbert space of gauge-invariant states is given by the configurations of integer-valued variables $\ell_j(\mathrm{r})$ on the links such that $\Delta_j \ell_j(\mathrm{r}) = 0$. In other words, we can think of the electric fields as a set of locally conserved, integer-valued "currents".

We can repeat the arguments we used in the $\mathbb{Z}_2$ gauge theory almost verbatim. The (trivially gauge-invariant) ground state, $|\text{gnd}_0\rangle$, is simply the state in which all the electric fields are zero on all links, $\ell_j(\boldsymbol{r}) = 0$, and the unperturbed ground state energy is zero, $E_{\text{gnd}_0} = 0$.

Since the electric fields are integer-valued, their sources must also be integers, $q(\boldsymbol{r}) \in \mathbb{Z}$. Hence, states created by a set of static sources $\{q(\boldsymbol{r})\}$, that we will label by $|\{q(\boldsymbol{r})\}\rangle$, obey a Gauss Law of the form

$$\Delta_j E_j(\boldsymbol{r})|\{q(\boldsymbol{r})\}\rangle = q(\boldsymbol{r})|\{q(\boldsymbol{r})\}\rangle \tag{18.62}$$

Let us consider now the case in which we have just two sources, with charges $\pm q \in \mathbb{Z}$, located at at sites $\boldsymbol{R}$ and $\boldsymbol{R}'$ of the lattice, just as we did in Fig.18.6a. The requirement of global charge neutrality follows from the fact that the global $U(1)$ symmetry is unbroken (it is easy to verify that the energy of states that violate global charge neutrality is divergent). The ground state in this sector of the Hilbert space is the state with $\ell_j(\boldsymbol{r}) = 0$ everywhere, except on the links on the shortest path, γ, stretching from $\boldsymbol{R}$ to $\boldsymbol{R}'$. Once again, the shortest path γ is just the set of links on the straight line between $\boldsymbol{R}$ and $\boldsymbol{R}'$. On each link on γ, the electric fields must take the value $\ell_j = q$. Therefore, the energy cost of this state with sources is, once again, a linear function of the separation of the sources, $V(|\boldsymbol{R} - \boldsymbol{R}'|) = \sigma|\boldsymbol{R} - \boldsymbol{R}'|$, where the string tension is $\sigma = \frac{q^2}{2} - O(1/g^2)$. Hence, this theory is confining in strong coupling regime. We will see in the next chapter that the $U(1)$ gauge theory has magnetic monopoles which play a key role in the confined phase.

Non-abelian Gauge Theory

This analysis applies to non-abelian Yang-Mills theories with minor, but important, changes. Indeed, in the strong coupling limit, the dominant term of the Hamiltonian of Eq.(18.27) is the kinetic energy term, which is proportional to the sum of the Casimir operators $E_j^2(\mathrm{r})$ on each link. Since the group in non-abelian, the eigenstates are those of the quadratic Casimir and the diagonal (Cartan) generators. For $SU(2)$, the states are $|j, m\rangle$ and carry the quantum numbers of the representations. The ground state is the $SU(2)$ singlet state $|0, 0\rangle$ on every link. This ground state energy of this trivially gauge-invariant state is zero. We can now consider the Hilbert space for a theory with two static sources, each must now carry $SU(2)$ quantum numbers, such that the total state remains a singlet. Again, this follows from the fact that the global $SU(2)$ symmetry is unbroken. For the case of two sources separated a distance R (in lattice units) that carry the spinor representation, $(1/2, 1/2)$, and its conjugate, one finds, once again, that the energy of this state follows a linear potential law, $V(R) = \sigma R$, where the string tension now is $\sigma = \frac{3}{8}g^2 - O(1/g^2)$. Hence, in the strong-coupling regime, the non-abelian theory is confining.

18.8 Hamiltonian Duality

Duality is a powerful tool to examine the topological properties of discrete gauge theories. We will now show that a theory with a $\mathbb{Z}_2$ symmetry (be global or local) has a *dual* theory defined on the dual lattice. This notion generalizes to theories with abelian symmetry groups. The main power of duality transformations is that, in general, relate weakly coupled theories to strongly coupled theories. In the spacial case of the symmetry group $\mathbb{Z}_2$, the dual has the same symmetry but is not necessarily realized in the same way. The extension of these concepts to more general theories, e.g. with non-abelian groups, is generally only possible, in supersymmetric theories. Modern String theory has greatly generalized the notion of duality accross theories and dimensions.

Conventionally, duality is an identifications of series expansions of two partition functions, and will be discussed briefly in the next chapter. The original notion of duality, and more specifically Kramers-Wannier duality, is a property of two-dimensional classical Ising models. It was subsequently extended by Wegner to $\mathbb{Z}_2$ gauge theory. Here we will use a Hamiltonian approach to duality. In this approach, we defined non-local operators, such as the kink operator of the $1+1$ dimensional quantum Ising model, discussed in Chapter 14, the $\mathbb{Z}_2$ monopole, and the magnetic ‘t Hooft operators, which do not commute with the local degrees of freedom.

The Hamiltonian of the one-dimensional quantum Ising model is

$$H_{\text{Ising}}(\lambda) = -\sum_{n=1}^{N} \sigma^1(n) - \lambda \sum_{n=1}^{N} \sigma^3(n)\sigma^3(n+1) \tag{18.63}$$

Here the lattice is a 1D chain. The dual lattice are the midpoints of the lattice, i.e. the set of points $\tilde{n} = n + 1/2$. The operators $\sigma^1(n)$ and $\sigma^3(n')$ satisfy the Pauli matrix (Clifford) algebra, i.e. they commute if $n \neq n'$, and anticommute if $n = n'$, and square to the identity. We will define a dual theory in terms of a new Clifford algebra defined on the sites of the dual lattice as

$$\tau^3(\tilde{n}) = \prod_{m\leq n} \sigma^1(m), \quad \tau^1(\tilde{n}) = \sigma^3(n)\sigma^3(n+1) \tag{18.64}$$

We recognize the operator $\tau^3(\tilde{n})$ as the kink creation operator defined in Chapter 14 (c.f. Eq.(14.51).) It is trivial to see that the operators of Eq.(18.64) obey the same Clifford algebra as the Pauli operators. Moreover, in terms of these operators the dual Hamiltonian takes the form (up to a term determined by the boundary conditions)

$$H_{\text{ising}}(\lambda) = -\sum_{\tilde{n}=1^N} \tau^3(\tilde{n})\tau^3(\tilde{n}+1) - \lambda \sum_{\tilde{n}=1}^{N} \tau^1(\tilde{n}) = \lambda \tilde{H}_{\text{Ising}}(1/\lambda) \tag{18.65}$$

Hence, the dual Hamiltonian, $\tilde{H}_{\text{Ising}}$, is the same as the original Hamiltonian up to a change $\tilde{\lambda} = 1/\lambda$. This mapping is the equivalent of the famous Kramers-Wannier self-duality of the 2D classical Ising model. We see that, in this theory the dual has the same $\mathbb{Z}_2$ global symmetry. Duality maps the weak coupling (disordered) phase to the strong coupling (broken symmetry) phase (and viceversa). The phase transition we discussed in Chapter 14 is at the self-dual point, $\lambda = 1$. We also see that the disordered phase can be viewed as condensate of kinks, i.e. $\langle \tau^3(\tilde{n}) \rangle \neq 0$ is the disordered phase of the theory.

Moving on to 2 + 1 dimensions we will consider the Ising gauge theory whose Hamiltonian is (c.f Eq.(18.21))

$$H_{\mathbb{Z}_2\mathrm{GT}} = -\sum_{\boldsymbol{r},j=1,2} \sigma_j^1(\boldsymbol{r}) - \lambda \sum_{\boldsymbol{r},j,k} \sigma_j^3(\boldsymbol{r})\sigma_k^3(\boldsymbol{r}+\boldsymbol{e}_j)\sigma_j^3(\boldsymbol{r}+\boldsymbol{e}_k)\sigma_k^3(\boldsymbol{r}) \tag{18.66}$$

The lattice now is a square lattice (the set of points labeled by $\boldsymbol{r}$), and the degrees of freedom (the gauge fields) are defined on the midpoints of the links of the lattice. As in the previous case, the operators obey a Clifford algebra, but the theory now has a local (gauge) $\mathbb{Z}_2$ symmetry. The dual of the square lattice is also a square lattice, the set of points on the center of the plaquettes of the square lattice, $\tilde{\boldsymbol{r}} = \boldsymbol{r} + (\boldsymbol{e}_1 + \boldsymbol{e}_2)/2$. We will define the operator $\tau^3(\tilde{\boldsymbol{r}})$ to be the $\mathbb{Z}_2$ monopole operator $M(\tilde{\boldsymbol{r}})$ of Eq.(18.58). Recall that the $\mathbb{Z}_2$ monopole operator is defined in terms of a Dirac string along an open path $\{\tilde{\gamma}$ of the dual lattice. We will define the operator $\tau^1(\tilde{\boldsymbol{r}})$ to be the plaquette operator, i.e. $\tau^1(\tilde{\boldsymbol{r}}) = \sigma_j^3(\boldsymbol{r})\sigma_k^3(\boldsymbol{r}+\boldsymbol{e}_j)\sigma_j^3(\boldsymbol{r}+\boldsymbol{e}_k)\sigma_k^3(\boldsymbol{r})$. It is straightforward to see that these operators obey the Clifford algebra since they anticommute on the same site of the dual lattice, commute otherwise, and square to the identity. Notice the important fact that the dual operators are defined on the sites of the dual lattice whereas the original degrees of freedom are defined on the links.

One can check that these definitions are consistent provided the gauge theory obeys the Gauss law condition, Eq.(18.53), which is trivially satisfied by these definitions. It is immediate to see that, up to boundary conditions, the Hamiltonian of the dual theory is that of the quantum Ising model in 2+1 dimensions!

$$H_{\mathbb{Z}_2\mathrm{GT}} = \lambda \tilde{H}_{\mathrm{Ising}}(1/\lambda) = -\lambda \sum_{\tilde{\boldsymbol{r}}} \tau^1(\tilde{\boldsymbol{r}}) - \sum_{\tilde{\boldsymbol{r}},j=1,2} \tau^3(\tilde{\boldsymbol{r}})\tau^3(\tilde{\boldsymbol{r}}+\boldsymbol{e}_j) \tag{18.67}$$

Clearly, the theory is no longer self dual. However, a generalization of this construction to the $\mathbb{Z}_2$ gauge theory with a $\mathbb{Z}_2$ matter field is actually self dual. Hence, in 2+1 dimensions duality also maps a weakly coupled theory to strongly coupled theory, and viceversa. But the theories are no longer the same as it now maps *the gauge-invariant sector* a theory with local symmetry, the $\mathbb{Z}_2$ gauge theory, to a theory with a global $\mathbb{Z}_2$ symmetry, the 2+1 dimensional quantum Ising model (equivalent to the three-dimensional classical Ising model.)

In particular, duality maps the broken symmetry phase of the Ising model, $\lambda < \lambda_c$, to the confined phase of the gauge theory, and that the order parameter of the Ising model maps onto the $\mathbb{Z}_2$ monopole operator of the gauge theory. Hence, as anticipated, the confining phase can be regarded as a condensate of $\mathbb{Z}_2$ magnetic monopoles. On the other hand, the disordered phase of the Ising model maps onto the deconfined phase of the $\mathbb{Z}_2$ gauge theory. We will see in the next section, that careful consideration of boundary conditions (which we sidestepped here) show that, in this phase, the gauge theory is a topological field theory. In fact, this is true in all dimensions.

Let us finally consider the $\mathbb{Z}_2$ gauge theory in 3+1 dimensions. As before, the degrees of freedom are defined on the links of the three dimensional cubic lattice, i.e. they are *vector fields*, and the $\mathbb{Z}_2$ flux is defined on the plaquettes of the lattice. In three dimensions plaquettes are oriented two-dimensional surfaces, Thus, as expected, the flux is an ani-symmetric tensor field. Now, in three dimensions, links are dual to plaquettes and viceversa. Therefore, the dual of the of the 3+1-dimensional $\mathbb{Z}_2$ gauge theory should also be a gauge theory and, as we will see with a $\mathbb{Z}_2$ gauge group.

Thus, we once again define a set of Pauli matrices $\tau^1_{\tilde{l}}(\tilde{\boldsymbol{r}})$ and $\tau^3_{\tilde{l}}(\tilde{\boldsymbol{r}})$ (where $\tilde{\boldsymbol{r}}$ is the site at the center of the cube) on the links of the dual lattice such that the dual of the plaquette $\boldsymbol{r}, jk$ is the link $(\tilde{\boldsymbol{r}}, \tilde{l})$, and we identify

$$\begin{aligned}
\tau^1_{\tilde{l}}(\tilde{\boldsymbol{r}}) =& \sigma^3_j(\boldsymbol{r})\sigma^3_k(\boldsymbol{r}+\boldsymbol{e}_j)\sigma^3_j(\boldsymbol{r}+\boldsymbol{e}_k)\sigma^3_k(\boldsymbol{r}), \\
\sigma^1_l(\boldsymbol{r}) =& \tau^3_{\tilde{j}}(\tilde{\boldsymbol{r}})\tau^3_{\tilde{k}}(\tilde{\boldsymbol{r}}+\boldsymbol{e}_{\tilde{j}})\tau^3_{\tilde{j}}(\tilde{\boldsymbol{r}}+\boldsymbol{e}_{\tilde{k}})\tau^3_{\tilde{k}}(\tilde{\boldsymbol{r}})
\end{aligned} \tag{18.68}$$

both of which, due to the $\mathbb{Z}_2$ Gauss law, satisfy the Bianchi identity that the product of the plaquette flux operators on each of the six the faces of the cube must be the identity operator. Hence, in 3+1 dimensions, the $\mathbb{Z}_2$ gauge theory is *self-dual*. Assuming that there is a unique phase transition, it must happen at the self-dual point, $\lambda_c = 1$. It is also straightforward to see that the dual of the Wilson loop operator is the the 't Hooft loop operator that we defined before. Here too, duality maps the gauge invariant sector of one theory at coupling λ to the gauge invariant sector of the dual at coupling $1/\lambda$, again up to boundary conditions.

We now see the pattern. Duality involves a geometric duality, essentially the duality of forms: in three dimensions sites are dual to volumes, and links are dual to plaquettes. Thus, in 3+1 dimensions, the dual of a gauge theory is a gauge theory. This also implies that the dual of a scalar is an antisymmetric tensor field (a Kalb-Ramond field), and that the dual of a the 3+1 dimensional Ising model is a theory of an antisymmetric tensor field defined on plaquettes.

However, why does the dual of a $\mathbb{Z}_2$ theory have the same symmetry group? It turns out that if the theory has an *abelian* symmetry group G, its representations are one-dimensional and form a group, the dual group $\tilde{G}$. For instance, the dual of the group $U(1)$ is the group of integers $\mathbb{Z}$. The group $\mathbb{Z}_2$ is special in that it has only two (one-dimensional) representations and their group is isomorphic to $\mathbb{Z}_2$. It is this feature of abelian groups the does not allow for generalizations to non abelian theories.

18.9 Confinement in the Euclidean Space-time Lattice Picture

Let us examine the behavior of gauge theory on a space-time Euclidean lattice in the strong coupling regime. The partition function of a gauge theory with gauge group G has the form (with a suitable gauge fixing condition)

$$Z[G] = \int \mathcal{D}\mathcal{U}_\mu \; \exp\left[-\frac{1}{g} \sum_{\text{plaquettes}} \left(\operatorname{tr} \mathcal{U}_\mu \mathcal{U}_\nu \mathcal{U}_\mu^{-1} \mathcal{U}_\nu^{-1} + \text{c.c.}\right)\right] \tag{18.69}$$

Here "plaquettes" means sum over all the plaquettes of the lattice.

As before, here the coupling constant g plays the role of temperature T in statistical mechanics. Therefore, we can examine the behavior of the partition function in the strong coupling regime by an analog of the high temperature expansion in statistical mechanics. In this limit, we have to expand the exponential in powers of $1/g$ and compute the averages (expectation values) over ll configurations of gauge fields, using as the only weight the Haar measure of the integration. Since the action is (by construction) gauge-invariant, gauge-equivalent configurations will yield the same value of the average. This amounts to a contribution to the partition function in the form of an overall factor of $v(G)^{ND}$, where $v(G)$ is the volume of the gauge group (which is finite for a compact group G), and N is the number of sites (the volume of spacetime) and D is the dimension. We recognize that it is this divergent contribution that is eliminated by a proper definition of the integration measure, which should sum over gauge classes and not configurations (see Chapter 9 and the Faddeev-Popov procedure).

To simplify the discussion (and the notation) we will describe the strong coupling expansion in the context of the $\mathbb{Z}_2$ gauge theory. The partition function for the $\mathbb{Z}_2$ gauge theory can be written as

$$
\begin{aligned}
Z[\mathbb{Z}_2] &= \mathrm{Tr}\exp\left(\sum_{\text{plaquettes}} \frac{1}{g}\tau_\mu\tau_\nu\tau_\mu\tau_\nu\right) \\
&= \left(\cosh\left(\frac{1}{g}\right)\right)^{ND(D-1)} \mathrm{Tr}\prod_{\text{plaquettes}}\left[1+\tau_\mu\tau_\nu\tau_\mu\tau_\nu \tanh\left(\frac{1}{g}\right)\right] \qquad (18.70)
\end{aligned}
$$

where the trace is over the $\mathbb{Z}_2$ gauge fields τ_μ at each link. For a general gauge group, the Gibbs weight for each plaquette is expanded as a sum of characters of the representations of the gauge group. The $\mathbb{Z}_2$ group has only two representations. In what follows, we will drop the overall prefactor since it does not contribute to expectation values.

Since $\mathrm{Tr}\tau_\mu = 0$, the only non-vanishing contributions must be such that each $\tau_\mu = \pm 1$ gauge field appears twice in the expansion. This means that each link must appear twice. Since each plaquette, by construction, only appears once, the non-vanishing contributions are sets of plaquettes such that all of their edges (links) are glued to other plaquettes in the set. In other words, the plaquettes in the set must cover closed surfaces. An example is presented in Fig.18.8, showing the first non-vanishing contribution to the partition function in $D = 3$ dimensions. It is the surface of a cube, and has six faces. Since each face contributes a factor of $\tanh(1/g)$, this set contributes a factor of $2\tanh^6(1/g)$. This representation holds in all dimensions.

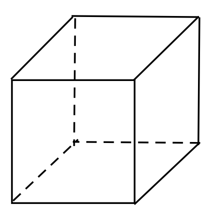

Figure 18.8 The leading non-vanishing contribution to the partition function of a $\mathbb{Z}_2$ gauge theory in $D = 3$ dimensions is the surface of a cube.

It follows that the partition function can be represented as a sum over sets of closed surfaces. In general, the non-vanishing terms of the strong coupling expansion is a sum over closed surfaces Σ of arbitrary genus $g = h-1$, here h is the number of handles of the surface. The sum has the form over surfaces of all possible genus g,

$$
Z = \sum_{\{\Sigma\}} (\tanh(1/g))^{\mathcal{A}[\Sigma]} \times \text{entropy factor} \qquad (18.71)
$$

where $\mathcal{A}[\Sigma]$ is the area of the surface Σ. The sum runs over all closed surfaces of any genus (with self-avoidance and non-overlapping conditions). The entropy factor counts the number of surfaces with the same area and genus.

A surface can be viewed as the history of a closed string. Fig.18.9a shows a closed surface Σ with genus 0 (i.e. with the topology of a sphere) as is being swept in imaginary time t by a string γ after being created at P_i and disappearing at P_f. Fig.18.9b shows the same process for a surface with the topology of a torus (genus 1). In this case, the closed string γ is created at P_i, splits into two closed strings γ_1 and γ_2 at Q_i, which rejoin at Q_f into the closed string γ, which, in turn, eventually disappears at P_f.

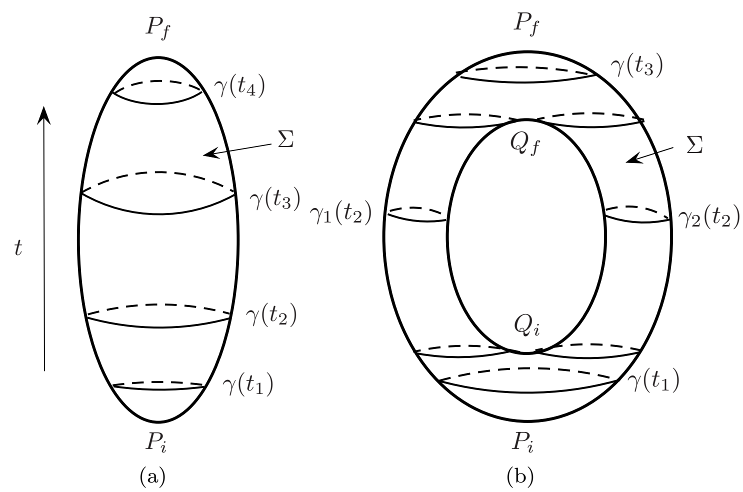

Figure 18.9 A closed surface Σ contributes to the strong coupling expansion by an amount $(2\tanh(1/g))^{A[\Sigma]}$, where $A[\Sigma]$ is its area (in lattice units). a) A closed surface with genus 0 (with the topology of a sphere); it can can be viewed as the history of a closed curve γ (a *string*) evolving in imaginary time t. As it sweeps the surface Σ. Here the closed string is created in the remote past at event P_i, evolves from times t_1 through t_4, and disappears at event P_f. b) A closed surface of genus 1 (a torus). Here the closed string γ is created at P_i, evolves to Q_i where it splits into two strings γ_1 and γ_2, which after evolving for some time, rejoin at Q_f into the single closed string γ, which eventually disappears at event P_f.

We will now turn to the behavior of the Wilson loop in the strong coupling limit, which we will see is drastically different from what we found in the weak coupling regime. The result that we will find in the strong coupling regime, holds for *all* gauge theories with a compact gauge group have the same behavior. We will see that the expectation value of the Wilson loop operator for a large loop is an *area law* exponential decay.

Before computing the expectation value of the Wilson loop operator we will use a rigorous result valid for any local theory that satisfies the condition of reflection positivity (i.e. Euclidean unitarity), discussed in Section 14.3, and applies to both theories with a global compact symmetry group and theories with a local compact symmetry group. In both types of theories, the Euclidean action is local and has the form $S = -\frac{1}{g}\mathcal{F}[\text{fields}]$. Then, in such theories the expectation values of correlators of observables are positive, $\langle \mathcal{O} \rangle \geq 0$, where $\mathcal{O}$ is an observable (or product of) invariant under the symmetry transformations. Moreover, the expectation values satisfy a monotonic dependence on the coupling constant g,

$$\langle \mathcal{O}[g_1] \rangle > \langle \mathcal{O}[g_2] \rangle, \qquad \text{for} \quad g_1 < g_2 \tag{18.72}$$

This inequality is an example of a Griffiths-Ginibre inequality.

In classical statistical mechanics, this inequality says the as the temperature T increases, the correlations decrease. This result holds for systems that obey reflection positivity such as ferromagnets (or, more generally, to unfrustrated magnets). In that context, it implies that the *fastest* possible decay of a correlation function at long distances is the exponential decay.

Since a gauge theory with a compact gauge group satisfies reflection positivity, hence the inequality of Eq.(18.72) also applies to these theories. In section 18.6 we saw that in a gauge theory, with a compact gauge group, only gauge-invariant observables have a non-vanishing expectation value. Of all such observables, the more relevant one is the Wilson loop operator, and this inequality applies to this operator. It implies that the expectation value of the Wilson loop operator at some coupling constant g_1 is bounded from below by its value at a larger value of the coupling constant g_2. We will now see that, for large enough coupling constant g, the expectation value of the Wilson loop decays as an exponential function of the are a bounded by the loop. Thus, in this regime it obeys an *area law*. It follows that the regimes where the Wilson loop obeys, respectively, area and perimeter laws must be separated by a phase transition.

We will now compute the expectation value of the Wilson loop operator, with the same geometry as shown in Fig.18.5. here we will do the computation for the simpler case of the $\mathbb{Z}_2$ gauge theory. However, the calculation works much in the same way for all compact gauge groups, both discrete and continuous. Using again a compact notation, we write the expectation value of the Wilson loop operator in the $\mathbb{Z}_2$ gauge theory as

$$\langle W_\gamma[\tau_\mu] \rangle_K = \frac{1}{Z[K]} \mathrm{Tr} \left[W_\gamma[\tau_\mu] \exp \left(K \sum_{\text{plaquettes}} K \tau_\mu \tau_\nu \tau_\mu \tau_\nu \right) \right] \tag{18.73}$$

where

$$W_\Gamma[\tau_\mu] = \prod_{(\mathrm{r},\mu)\in\Gamma} \tau_\mu(\mathrm{r}) \tag{18.74}$$

is the Wilson loop operator.

The computation is elementary. In the strong coupling regime, $K = \frac{1}{g} \ll 1$, we expand the exponential in powers of $\tanh(K)$. The only non-vanishing terms in the trace are such that the $\mathbb{Z}_2$ gauge fields on Γ are matched by gauge fields from the plaquettes in the exponential. The only surviving terms must be such that all $\mathbb{Z}_2$ gauge fields of γ and from the set of plaquettes are matched. It is easy to see that these contributions correspond to the tiling of open surfaces Σ whose boundary is Γ. Each term of the sum contributes with a factor of $(2\tanh K)^{\mathcal{A}[\Sigma]}$ times a multiplicity factor. It follows that the leading non-vanishing term is the tiling of the minimal surface Σ_{minimal} with boundary Γ, shown in Fig.18.10. In the case of a planar loop there is

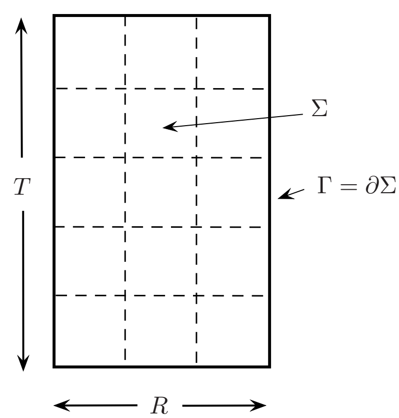

Figure 18.10 The leading contribution to the expectation value of the Wilson loop on Γ is the minimal surface Σ.

only one minimal surface and there is no multiplicity factor. In this case, the area of the minimal surface is equal to the number of plaquettes enclosed by the loop. For a loop of (imaginary) time span T and spatial extent R (both in lattice units), the area enclosed by the loop is $\mathcal{A}[\Sigma] = RT$. Therefore, at leading order, we find that the expectation value of the Wilson loop decays exponentially with the area enclosed by the loop as

$$\langle W_\Gamma \rangle_K = \exp\left(-\sigma[K]\mathcal{A}[\Sigma]\right) \tag{18.75}$$

where, at leading order, the *string tension* $\sigma(K)$ is

$$\sigma(K) = -\ln\tanh K + \ldots \simeq \ln g + \ldots \tag{18.76}$$

This result, combined with the Griffiths-Ginibre Inequality, Eq.(18.72), proves that the expectation value of the Wilson loop in a gauge theory with a compact gauge group is bounded by an function that decays exponentially for large loops following an *area law*. Hence, the Area Law represents the *fastest* rate at which a Wilson loop can decay for loops of large sizes.

We can now readily obtain the effective potential $V(R)$ to be

$$\begin{aligned} V(R) &= -\lim_{T/R\to\infty} \ln\langle W_\Gamma\rangle_K \\ &= \sigma(K)R \end{aligned} \tag{18.77}$$

where the string tension, $\sigma(K)$, is given in Eq.(18.76).

We found that the effective potential $V(R)$ grows linearly with separation. This implies that the $\mathbb{Z}_2$ sources introduced by the Wilson loop, are *confined* since the cost of separating them all the way to infinity is divergent. In contrast, in the $g \ll 1$ regime, we found that the energy cost is finite, the self-energy of the sources.

18.10 Behavior of Gauge Theories coupled to Matter Fields

In the absence of matter fields, gauge theories have two possible *phases*: a) a deconfined phase, in which the Wilson loop has a Perimeter Law (or, possibly, Coulomb behavior), and the effective interaction between external charges (sources) has an exponential decay (or a $1/R$ fall off), and b) a confined phase, in which the Wilson loop has an Area Law and the interaction between static sources grows linearly with distance. Confined and deconfined phases are separated by phase transitions, which may be continuous or discrete (first order). If the phase transitions are continuous, they define a non trivial continuum quantum field theory.

In the case of gauge theories with a discrete gauge group, they may have a deconfined phase if the space-time dimension $D > D_c = 2$, which is the lower critical dimension. Instead, the lowest critical dimension for a gauge theory with a compact continuous gauge group to have a deconfined phase is $D_c = 4$ space-time dimensions, which is also the dimension in which their coupling constant is dimensionless and the theory is renormalizable in perturbation theory. This behavior is in close similarity to what we found for matter fields: matter fields with a discrete global symmetry, e.g. the Ising model, have a lower critical dimension $D_c = 1$ above which the symmetry may be spontaneously broken, and the lower critical dimension for matter fields with a continuous global symmetry is $D_c = 2$ (again, where the theories are perturbatively renormalizable).

We will now ask what happens if the gauge fields are (minimally) coupled to dynamical matter fields. In our discussion matter fields have only entered as static sources, i.e. as infinitely heavy particles that carry the gauge charge. Clearly, if the matter fields are sufficiently heavy (and "uncondensed"), the expectation is that their effects should be mild, and amount, e.g. to a renormalization of the coupling constant of the gauge fields.

However, the behavior of some observables change, no matter how heavy the matter fields are. Consider a gauge theory deep in its confining phase. We saw that, when the matter fields are absent or, rather, are static sources, that the Wilson loop operator has an area law, and that the energy to separate two charges grows linearly with their separation, as shown in Fig.18.6b.

We will now see that, if the dynamical matter fields carry the fundamental charge of the gauge group, for large enough Wilson loops, the area law behavior always yields to a perimeter law behavior. In spite of that, we will see that the theory is still confining.

We will consider first the case of a $\mathbb{Z}_2$ gauge theory coupled to Ising matter field. We will work with the space-time lattice picture. The partition function for this theory is (using once again a synthetic notation)

$$Z[\beta, K] = \sum_{\{\tau_\mu\},\{\sigma\}} \exp\left(\beta \sum_{\text{links}} \sigma\tau_\mu\sigma + K \sum_{\text{plaquettes}} \tau_\mu\tau_\nu\tau_\mu\tau_\nu\right) \tag{18.78}$$

We will compute the expectation value of the Wilson loop operator on a closed contour Γ, as defined in Eq.(18.74). Notice that, since the $\mathbb{Z}_2$ gauge group has only one non-trivial representation, the Wilson loop has to carry the charge of the representation. So, in this case, there is only one Wilson loop that can be defined.

We will be interested in the regime where confinement is strongest, i.e. $K \ll 1$, and the $\mathbb{Z}_2$ matter field is heavy (and uncondensed), i.e. $\beta \ll 1$. In this regime, we can expand the Gibbs weight of this partition function in powers of K and β, and determine what is the leading contribution to the expectation value of the Wilson loop. At $\beta = 0$ we found that in the leading contribution we had to tile the Wilson loop with plaquettes spanning a minimal surface bounded by the contour Γ. This gave an area law contribution.

We can similarly set $K = 0$ ("ultra confinement") and seek the leading order contribution in β. In the definition of the Wilson loop, each link that belongs to Γ has a gauge field variable τ_μ which get multiplied around the loop. The leading term that cancels these link variables is the product of the link terms of the gauge-matter term of the action on the links that belong to Γ. In this product, each matter (Ising) variable appears twice (and squares to 1) leaving behind the product of the gauge variables on the links. These, in turn, will square to unity when multiplied by the gauge field on the Wilson loop. Therefore, for any value of β, no matter how small, there will be a *perimeter* law contribution.

Thus, in this regime, the lowest order contributions to the Wilson loop are

$$\langle W_\Gamma \rangle_{K,\beta} = (\tanh K)^{RT} + (\tanh \beta)^{2(R+T)} + \dots \tag{18.79}$$

where we used that the area of the minimal surface is $A = RT$ and the perimeter of the loop is $L = 2(R+T)$. Clearly, for large enough loops, the perimeter law contribution wins over the area law. For loops with $T \gg R \gg 1$, effective potential now will be

$$V(R) = \begin{cases} \sigma(K)\, R, & \text{for } R \ll \xi \\ 2\rho(k), & \text{for } R \gg \xi \end{cases} \tag{18.80}$$

where, at this order, $\sigma(K) = -\ln \tanh K + \dots$ is the string tension, $\rho(k) = -\ln \tanh \beta + \dots$ is the mass (self-energy) of the matter field, and the crossover scale between the two behaviors is $\xi = \sigma(K)/\rho(K)$.

Clearly, this implies that we have linear potential until a scale $\xi \simeq 2\rho/\sigma$, where the string breaks by creating a pair of excitations of the matter field (from the vacuum), which screen that test charges of the Wilson loop. Beyond this scale, the potential essentially the constant value $V_{\max}$, equal to twice the mass of the matter field (see. Fig.18.11). Physically, the saturation of the linear potential means that the test charges (the Wilson loop) are being exactly screened (or compensated) by the creation of a pair of excitations from the vacuum. This behavior does not imply that confinement breaks down, since this excitation is a gauge-invariant state made of two $\mathbb{Z}_2$ matter fields stretching a string of the $\mathbb{Z}_2$ gauge field (see Fig.18.12.) Furthermore, in this phase the only local excitations are created by an operator $P(\text{plaquette})$ on small loops of gauge fields around a plaquette (i.e. a "glueball") separated a distance R, and tight bound states $\pi(\text{link})$ of matter and gauge fields ("pions"), also separated at a distance R. An elementary

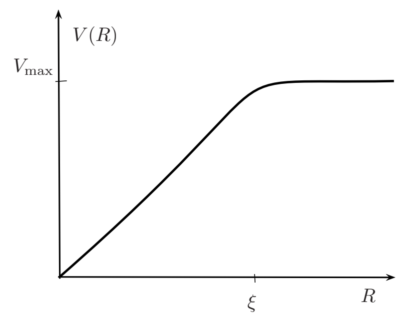

Figure 18.11 The effective potential for the $\mathbb{Z}_2$ gauge theory with dynamical $\mathbb{Z}_2$ (Ising) matter.

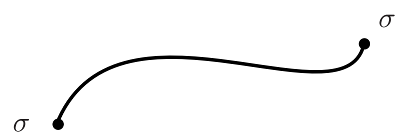

Figure 18.12 The effective potential for the $\mathbb{Z}_2$ gauge theory with dynamical $\mathbb{Z}_2$ (Ising) matter.

calculation shows that the (connected) correlator of two plaquette operators has the behavior

$$\langle P(\text{plaquette})P(\text{plaquette}')\rangle_c \simeq \exp(-M_P R) \tag{18.81}$$

where the mass of this excitation is $M_P = 4|\ln\tanh K|$. Likewise, the (connected) correlator of two "pions" on links separated a distance R, in this regime is

$$\langle \pi(\text{link})\pi(\text{link}')\rangle_c \simeq \exp(-M_\pi R) \tag{18.82}$$

where, at this level of approximation, the mass is $M_\pi = 2|\ln\tanh\beta| + |\ln\tanh K|$.

Hence, in spite of the perimeter-law behavior, all states are local gauge-invariant states, as required by Elitzur's theorem. Furthermore, the spectrum consists of massive locally gauge-invariant bound states. It is straightforward to show that this behavior holds for any theory with any compact gauge group in a confining phase. In addition, in the confining phase of the $\mathbb{Z}_2$ theory, there are no states in the spectrum that carry the $\mathbb{Z}_2$ charge. We will see shortly that, in the $\mathbb{Z}_2$ theory, such states do exist in the deconfined phase, although the associated operator is non-local. This behavior turns out to be due to the fact that in the deconfined phase the $\mathbb{Z}_2$ theory is topological. Similar behavior will be found in other theories with a discrete gauge group, e.g. $\mathbb{Z}_N$.

The "compensation" of the test charge that we found should not be confused with more conventional Debye screening. We will call this behavior, "algebraic screening". To understand the difference we need to consider a theory in which more than one charge is possible. One simple example is a theory with gauge group $U(1)$. The partition function now is

$$Z[\beta, K] = \int_0^{2\pi} \frac{d\theta}{2\pi} \int_0^{2\pi} \frac{dA_\mu}{2\pi} \exp\Big(\beta \sum_{\text{links}} \cos(\Delta_\mu \theta - qA_\mu) + K \sum_{\text{plaquettes}} \cos F_{\mu\nu}\Big) \tag{18.83}$$

where we denoted the flux on a plaquette by $F_{\mu\nu} = \Delta_\mu A_\nu - \Delta_\nu A_\mu$, and $q \in \mathbb{Z}$ is the $U(1)$ charge of the matter field $\exp(i\theta)$. The symmetry allows us to expand the Gibbs weights in the representations of the group $U(1)$:

$$\exp(\beta \cos\theta) = \sum_{\ell \in \mathbb{Z}} I_\ell[\beta] \exp(i\ell\theta) \tag{18.84}$$

This Fourier expansion can be regarded as an expansion in the characters of the representations of the group $U(1)$, which are just $\chi_\ell(\theta) = \exp(i\ell\theta)$, where $\ell \in \mathbb{Z}$. The character expansion applies to any group. In the specific case of the $U(1)$ group, the coefficients of this expansion are the modified Bessel functions $I_\ell(\beta)$, given by

$$I_\ell(\beta) = I_{-\ell}(\beta) \int_0^{2\pi} \frac{d\theta}{2\pi} e^{\beta\cos\theta} e^{-i\ell\theta} \tag{18.85}$$

For $\beta \ll 1$, it is well approximated by $I_\ell(\beta) \simeq (\beta/2)^\ell / \ell!$.

Using these expansions, the partition function of the $U(1)$ theory, of Eq.(18.83), becomes

$$Z[\beta, K] = \sum_{\{\ell_\mu\},\{m_{\mu\nu}\}} \prod_{\text{sites}} \delta_{\Delta_\mu \ell_\mu, 0} \prod_{\text{links}} \delta_{\Delta_\nu m_{\mu\nu} + q\ell_\mu, 0} \prod_{\text{links}} I_{\ell_\mu}[\beta] \prod_{\text{plaquettes}} I_{m_{\mu\nu}}[K] \tag{18.86}$$

where the variables $\ell_\mu \in \mathbb{Z}$ and $m_{\mu\nu} \in \mathbb{Z}$ are defined, respectively, on the links and plaquettes of the lattice.

The Wilson loop operator with test charge $p \in \mathbb{Z}$, on a closed loop of the lattice Γ is

$$W_p[\Gamma] = \exp\left(ip \sum_{(\boldsymbol{r},\mu)\in\Gamma} s_\mu(\boldsymbol{r}) A_\mu(\boldsymbol{r}) \right) \tag{18.87}$$

where $s_\mu(\boldsymbol{r})$ is an oriented kink variable that is $s_\mu = 1$ on links on the closed contour Γ, zero otherwise. The expectation value of this operator is computed by making the following replacement in the partition function of Eq.(18.86):

$$\prod_{\text{links}} \delta_{\Delta_\nu m_{\mu\nu}+q\ell_\mu,0} \mapsto \prod_{\text{links}} \delta_{\Delta_\nu m_{\mu\nu}+q\ell_\mu,ps_\mu} \tag{18.88}$$

Deep in the confined phase, $K \ll 1$ and $\beta = 0$, the link variables are set to zero, $\ell_\mu = 0$. In this regime, we get the area law behavior for Wilson loops with any test charge p. In this limit, only configurations with $m_{\mu\nu} = p$ on the smallest possible number of plaquettes can contribute, i.e. plaquettes $(\boldsymbol{r}, \mu\nu)$ that tile the minimal surface Σ whose boundary is the contour Γ. Hence, the leading configuration is $m_{\mu\nu}(\boldsymbol{r}) = p\Theta(\boldsymbol{r}, \mu\nu)$, where $\Theta(\boldsymbol{r}, \mu\nu) = 1$ if the plaquette $(\boldsymbol{r}, \mu\nu) \in \Sigma$, and zero otherwise. Their total contribution to expectation value of the Wilson loop of charge p is an area law

$$\langle W_p[\Gamma]\rangle = \left(\frac{I_p(K)}{I_0(K)}\right)^{RT} \tag{18.89}$$

and we find a potential linear in R, with a string tension

$$\sigma(K,p) = -\ln(I_p(K)/I_0(K)) \approx \ln(p!(2/K)^p) + \dots, \tag{18.90}$$

which is finite for all p, and, in this limit, is large.

However, for $\beta \ll 1$ but finite, we need to satisfy two constraints, $\Delta_\mu \ell_\mu = 0$ and $\Delta_\nu m_{\mu\nu} + q\ell_\mu = ps_\mu$, while making both $|m_{\mu\nu}|$ and $|\ell_\mu|$ as small as possible. We find several behaviors. If the charge of the Wilson loop is a multiple of the charge of the dynamical matter field, $p = qr$ (for some integer r), we find two leading contributions: a) $\ell_\mu = 0$ and $m_{\mu\nu} = p$ on the plaquettes that tile the minimal surface Σ, and b) $m_{\mu\nu} = 0$ everywhere and $\ell_\mu = r$ on links on the contour Γ. The first contribution reproduces the area law behavior, while the second yields a perimeter law

$$\langle W_{p=qr}[\Gamma]\rangle = \left(\frac{I_p(K)}{I_0(K)}\right)^{RT} + \left(\frac{I_r(\beta)}{I_0(\beta)}\right)^{2(R+T)} \tag{18.91}$$

Hence, for $p = qr$, the interaction potential saturates to the value $V_{\max} = -2\ln(I_r(\beta)/I_0(\beta)) \approx 2\ln(r!(2/\beta)^r) + \dots$, and, for sufficiently large loops, we find the same algebraic screening we found before.

On the other hand, if the charge of the Wilson loop is smaller than the charge of the dynamical matter field, $p < q$, then the leading contribution comes form configurations with $\ell_\mu = 0$ and $m_{\mu\nu} = p$ on the plaquettes of the minimal surface (and zero otherwise). In other words, in this case the leading behavior is the area law of Eq.(18.89), the effective potential is linear in R, and there is no algebraic screening.

Similarly, if the charges are not multiples of each other, but still $p > q$

(e.g. $p = qr + k$), the solutions to the constrains now are $\ell_\mu = qs_\mu$ (only on the links of Γ), and $m_{\mu\nu} = k$ only on the plaquettes of the minimal surface bounded by Γ. Thus, we now find

$$\langle W_{p=qr+k}[\Gamma]\rangle = \left(\frac{I_k(K)}{I_0(K)}\right)^{RT}\left(\frac{I_r(\beta)}{I_0(\beta)}\right)^{2(R+T)} \tag{18.92}$$

In other words, we recover an area law even if the matter field is dynamical, and the effective potential no longer saturates. Now, the unscreened part of the test charge k sees a linear potential with a non-vanishing (but smaller) string tension $\sigma(K,k) = \ln(k!(2/K)^k) + \ldots$, and an offset of the linear potential, $V_{\text{off}} = 2\ln(r!(2/\beta)^r)$.

These results, which represent the leading behavior of a convergent expansion in powers of K and β, show that when to charge of the test charge and that of the dynamical matter field are not proportional to each other, the Wilson loop of the test charge has an area law behavior. Only when proportionality holds we get a perimeter law and saturation, and hence, algebraic screening. The same behavior is found in more general cases, e.g. in a non-abelian gauge theory with matter in the adjoint representation, the Wilson loop for test charge in the fundamental representation will exhibit confinement.

18.11 The Higgs Mechanism

The Higgs mechanism arises in theories of gauge and matter fields with a continuous symmetry in the regime where the matter field spontaneously breaks the global symmetry.

18.11.1 The Abelian Higgs Model

In the simplest case the symmetry group is $U(1)$, and the matter field is a complex scalar field ϕ coupled to the Maxwell's electrodynamics. The Lagrangian density (in Minkowski space-time) of this theory, known as the abelian Higgs model, is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + |D_\mu\phi|^2 - V(|\phi|^2) \tag{18.93}$$

where $D_\mu = \partial_\mu + ieA_\mu$ is the covariant derivative. The potential is, as usual, $V(|\phi|^2) = m^2|\phi|^2 + \lambda|\phi|^4$. If $m^2 < 0$, classically it has a minimum at $|\phi_0| = (|m^2|/3\lambda)^{1/2}$, and the global $U(1)$ symmetry is spontaneously broken.

In the Euclidean domain, the Lagrangian becomes

$$\mathcal{L}_E = |D_\mu \phi|^2 + m^2 |\phi|^2 + \lambda |\phi|^4 + \frac{1}{4} F_{\mu\nu}^2 \tag{18.94}$$

which is identical to the Landau-Ginzburg theory for a superconductor, with ϕ being the order parameter field of the superconductor (the Cooper pair condensate), and the charge is replaced by $2e$. The path integral of this model describes the classical partition function of a superconductor interacting withe the thermal fluctuations of the magnetic field.

Let us rewrite the complex scalar field in terms of an amplitude and a phase field,

$$\phi = \rho e^{i\theta} \tag{18.95}$$

We will focus on the phase, $m^2 < 0$, in which the scalar field has a broken (global) $U(1)$ symmetry. In this phase, the phase field θ is the Goldstone boson of the spontaneously broken global $U(1)$ symmetry. Deep in this phase, the amplitude field is essentially pinned to its classical expectation value, $\rho_0 = |\phi_0|$. Then, the Lagrangian of the abelian Higgs model, Eq.(18.93), becomes

$$\mathcal{L} = \frac{\rho_0^2}{2} \left(\partial_\mu \theta + e A_\mu \right)^2 - \frac{1}{4} F_{\mu\nu}^2 \tag{18.96}$$

We can now fix the gauge $\theta = 0$, known as the unitary gauge (the same as the London gauge in superconductivity), or, equivalently, make a gauge transformation, $A_\mu \to A_\mu - \partial_\mu \Phi$, with $\theta = -e\Phi$. In both descriptions, the phase field θ disappears from the theory, i.e. the Goldstone boson is "eaten" by the gauge field. The Lagrangian now is

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 + \frac{1}{2} e^2 \rho_0^2 A_\mu^2 \tag{18.97}$$

Hence, in the broken symmetry phase, the gauge field becomes massive, and in this theory the mass of the photon is $m = e\langle\phi\rangle$. The phenomenon of the gauge field becoming massive upon eating a Goldstone boson is known as the Higgs mechanism. Physically, this phenomenon is equivalent to the expulsion of a magnetic flux in a superconductor, known as the Meissner effect.

Another key feature of the Higgs fields is that they yield masses to Dirac fermions. Thus, consider adding a massless Dirac field ψ to this theory, minimally coupled to the gauge field A_μ, and that the fermions also couple to the Higgs field through Yukawa couplings. The fermionic sector of the theory now is

$$\mathcal{L}_{\text{fermions}} = \bar{\psi} i \not{D} \psi + G \bar{\psi} \psi \phi \tag{18.98}$$

We now see that, if the scalar field has a vacuum expectation value, $v = \langle\phi\rangle$, the Yukawa coupling becomes a fermion mass term with a mass $m = Gv$, or equivalently, that $\langle\bar{\psi}\psi\rangle = G\langle\phi\rangle$. The reader familiar with the BCS theory of superconductivity will recognize that there is a very close parallel between the mechanism of mass generation via the Higgs mechanism and the development of a superconducting gap by Cooper-pair condensation.

The analysis that we have done here is classical, and the computation of perturbative corrections are largely straightforward, except for the theory of the phase transition from the symmetric to the Higgs phase. However, the non-perturbative behavior, that we will discuss it shortly, is subtle, as we shall see.

18.11.2 The Georgi-Glashow Model

The Higgs mechanism plays a key role in the theory of weak interactions. An example is the Georgi-Glashow model which is a theory that unifies the weak and strong interactions with electromagnetism (Georgi and Glashow, 1974).

This theory involves a three-component real scalar field $\boldsymbol{\phi}$, i.e. a Higgs field with a global $O(3)$ spontaneously broken symmetry, coupled to an $SU(2)$ gauge field, $\boldsymbol{A}_\mu$, which is a matrix with values in the algebra of the group $SU(2)$, i.e. $\boldsymbol{A}_\mu = A_\mu^a t^a$, where t^a are the three generators of $SU(2)$. The Lagrangian of this theory in the broken symmetry state is (in Euclidean space-time)

$$\mathcal{L} = \frac{1}{2}\left(D_\mu\boldsymbol{\phi}\right)^2 - \frac{1}{2}m^2\boldsymbol{\phi}^2 + \frac{\lambda}{4!}(\boldsymbol{\phi}^2)^2 + \frac{1}{4g^2}\mathrm{tr}F_{\mu\nu}^2 \tag{18.99}$$

where, for the gauge group $SU(2)$, the covariant derivative is

$$D_\mu\boldsymbol{\phi} = \partial_\mu\boldsymbol{\phi} + \boldsymbol{A}_\mu \times \boldsymbol{\phi} \tag{18.100}$$

and the field strength is

$$F_{\mu\nu} = i[D_\mu, D_\nu] = \partial_\mu\boldsymbol{A}_\nu - \partial_\nu\boldsymbol{A}_\mu + \boldsymbol{A}_\mu \times \boldsymbol{A}_\nu \tag{18.101}$$

We will write the scalar field as $\boldsymbol{\phi}^T = (\phi_1, \phi_2, \phi_3)$, and assume that the pattern of spontaneous symmetry breaking is $\boldsymbol{\phi} = (0, 0, m/\sqrt{\lambda})$. This classical expectation value breaks the $O(3)$ symmetry down to $U(1)$. By repeating the line of reasoning that we used in the abelian Higgs model, we find that in the broken symmetry state, the triplet of gauge fields of the Georgi-Glashow model can be rearranged as

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(A_\mu^{(1)} \mp iA_\mu^{(2)}), \quad A_\mu = A_\mu^{(3)} \tag{18.102}$$

It is straightforward to see that in this theory the Higgs mechanism implies that the doublet $W_\mu^\pm$ has a mass (squared) $m_W^2 = g^2 m^2/\lambda$, and that the field $A_\mu^{(3)}$ is massless. The conclusion is that the massive gauge bosons $W_\mu^\pm$ mediate the weak interactions, while the massless gauge fields $A_\mu^{(3)}$ are the photons of the electromagnetic sector. Also, just as in the case of the abelian Higgs model, the two Goldstone bosons of the spontaneously broken $O(3)$ symmetry disappear from the spectrum. On the other hand, the the longitudinal component, $\sigma = \phi_3 - m/\sqrt{\lambda}$, of the field $\boldsymbol{\phi}$ is massive, with $m_\sigma^2 = 2m_0^2$. This massive excitation is the Higgs particle.

The full theory of weak interactions, the Weinberg-Salam model, is an extension of the Georgi-Glashow model which includes, among other fields, $SU(2)$ doublets of Dirac fermions (electrons and neutrinos). In this theory, the fermions acquire a mass through Yukawa couplings of the fermions with the Higgs fields.

18.11.3 Observables of the Higgs Phase

It is tempting to regard the complex scalar fields of the Higgs models as order parameters, as they are if the symmetry is global. However, in a theory in which the symmetry is local, i.e. it is a gauge symmetry, the complex scalar field $\phi(x)$ (and its generalizations) is not gauge-invariant. This naturally leads to the question, what is the gauge-invariant meaning of the concept of spontaneous symmetry breaking when the symmetry is local? In other words, can we make the concept of spontaneous breaking of a local symmetry compatible with the requirements of Elitzur's theorem? For the concept of spontaneous symmetry breaking of a local symmetry to be meaningful, it has to be possible to construct a gauge-invariant order parameter that uniquely distinguishes the Higgs phase. Otherwise, the Higgs phase would not be a real phase of the theory.

Let us discuss this first in the case of a theory in which the gauge group is $\mathbb{R}$, and the theory has a Maxwell gauge field A_μ. A simple way to make the correlators of the Higgs field $\phi(x)$ gauge invariant is to define the non-local operator

$$G_\Gamma(x,y) = \left\langle \phi^\dagger(x) \exp\left(ie \int_{\Gamma(x,y)} dz_\mu A^\mu(z) \right) \phi(y) \right\rangle \tag{18.103}$$

where $\Gamma(x, y)$ is an arbitrary path stretching from x to y. This correlator is trivially invariant under the local gauge transformations $A_\mu \to A_\mu + \partial_\mu \Phi$ and $\phi(x) \to e^{ie\Phi(x)}\phi(x)$. Deep in the broken symmetry state, $m^2 < 0$, we can approximate $\phi(x) \simeq \phi_0 e^{i\theta(x)}$, where ϕ_0 is real and constant. In the unitary (London) gauge, $\theta = 0$, the correlator reduces to the computation of the expectation value

$$G_\Gamma(x, y) \simeq \phi_0^2 \left\langle \exp\left(ie \int_{\Gamma(x,y)} dz_\mu A^\mu(z)\right)\right\rangle \tag{18.104}$$

Since the gauge field is massive, it is easy to check that this correlator decays exponentially fast for large separations $|x - y|$. Therefore, this gauge-invariant operator is not an order parameter of the spontaneously broken symmetry but a test of the massive gauge field. In fact, this path-dependent operator can be defined for any gauge theory, both abelian and non-abelian.

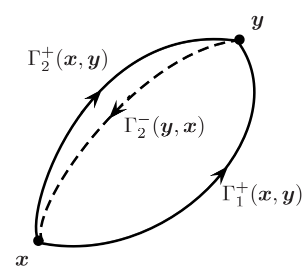

Figure 18.13 Path-dependence of the gauge-invariant correlator of Eq.(18.103). Here $\Gamma_1^+(x, y)$ and $\Gamma_2^+(x, y)$ are two oriented paths stretching from x to y. $\Gamma_2^-(y, x)$ is the reversed path $\Gamma_2^+(x, y)$.

Moreover, the non-local gauge-invariant correlator of Eq.(18.103) is *path-dependent* (as shown in Fig.18.13):

$$\begin{aligned} \exp\left(ie \int_{\Gamma_1^+} dz_\mu A^\mu(z)\right) &= \exp\left(ie \int_{\Gamma_2^+} dz_\mu A^\mu(z)\right) \times \exp\left(ie \oint_{\Gamma_1^+ \cup \Gamma_2^-} dz_\mu A^\mu(z)\right) \\ &= \exp\left(ie \int_{\Gamma_2^+} dz_\mu A^\mu(z)\right) \times \exp\left(i\frac{e}{2} \int_\Sigma dS_{\mu\nu} F^{\mu\nu}\right) \end{aligned} \tag{18.105}$$

where $\partial\Sigma = \Gamma_1^+ \cup \Gamma_2^-$. Clearly, the operator depends on the amount of flux in the surface Σ bounded by the two paths. Although the factorization that we used above holds only for abelian theories, the path-dependence of the operator also holds for non-abelian theories.

Is there an alternative? Yes, there is provided the gauge group is $\mathbb{R}$, i.e. a non-compact gauge theory. In fact, in the context of quantum electrodynamics, Dirac observed that it is unphysical (and violates gauge invariance) to define an operator that creates a charged particle without its static Coulomb field, i.e. without creating a coherent state of photons. This is done by considering instead the non-local operator

$$\phi^\dagger(\boldsymbol{x}) \exp\left(ie \int d^3z \boldsymbol{E}(\boldsymbol{z}) \cdot \boldsymbol{A}(\boldsymbol{z}) \right) \phi(\boldsymbol{y}) \tag{18.106}$$

where the integral in the exponential extends over all space, and $\boldsymbol{E}(\boldsymbol{z})$ is the classical electric (Coulomb) field created by the charges at $\boldsymbol{x}$ and $\boldsymbol{y}$, and $\boldsymbol{A}$ is the vector potential of the theory. Under a gauge transformation both the operators $\phi^\dagger(x)$ and $\phi(y)$, and the exponential operator (that creates a coherent state of photons) change under a gauge transformation as

$$\phi(x) \mapsto \phi(x) e^{ie\Phi(x)}, \quad \boldsymbol{A}(x) \mapsto \boldsymbol{A}(x) + \nabla \Phi(x) \tag{18.107}$$

where $\Phi(z) \to 0$ as $|z| \to \infty$. However, gauge-invariance is maintained if the gauge-dependent contributions cancel each other out, i.e. if

$$e^{-ie\Phi(x)} e^{ie\Phi(y)} e^{i \int d^3z \boldsymbol{E}(z) \cdot \nabla \Phi(z)} = 1 \tag{18.108}$$

This condition can be met if the vector field $\boldsymbol{E}(z)$ satisfies

$$\nabla \cdot \boldsymbol{E}(z) = \rho(z) \equiv e\delta^3(z-x) - e\delta^3(z-y) \tag{18.109}$$

In other words, the vector field $\boldsymbol{E}(z)$ is just the classical Coulomb (electric) field created by the two charges $\pm e$ located at x and y, respectively. Furthermore, the operator

$$\exp\left(ie \int d^3z\, \boldsymbol{E}(z) \cdot \boldsymbol{A}(z) \right) \tag{18.110}$$

is a coherent state of photons. Also, since we can always solve the Poisson equation, Eq.(18.109), in terms of the potential $U(z)$ such that $\boldsymbol{E} = -\nabla U$, with $-\nabla^2 U = \rho$, we see that exponential in the coherent state operator becomes

$$\begin{aligned} \int d^3z\, \boldsymbol{E}(z) \cdot \boldsymbol{A}(z) &= - \int d^3z\, \nabla U(z) \cdot \boldsymbol{A}(z) \\ &= \int d^3z\, U(z) \nabla \cdot \boldsymbol{A}(z) \end{aligned} \tag{18.111}$$

In particular, *in the Coulomb* gauge, $\nabla \cdot \boldsymbol{A}(z) = 0$, the coherent state operator reduces to the identity operator. Hence, provided the Coulomb gauge can be defined unambiguously, the non-local operator of Eq.(18.106) reduces to the correlation function of the complex scalar fields.

Can we always define the operator of Eq.(18.106)? The answer to this question is that this is possible only if the Coulomb gauge is well defined. However, this is only possible for the Maxwell theory since it is abelian and non-compact. In fact, there is a topological obstruction to the definition of the Coulomb gauge if the group is compact (abelian and non-abelian). We will see in the next chapter that this obstruction has a topological origin and that it is related to the existence of magnetic monopoles. Therefore, we conclude that for a general compact gauge group it is not possible to construct an operator that has the properties of an order parameter.

18.12 Phase Diagrams of Gauge-Matter Theories

We just showed that in a general compact gauge theory it is not possible to find a gauge-invariant observable related to the order parameter $\langle \phi \rangle$ of the theory with a spontaneously broken *global* symmetry. These results naturally raise several questions

1. For a compact gauge group, can the Higgs phase be stable non-perturbatively? In other words, is there a Higgs phase distinct from the other phases of a gauge theory?
2. What is the relation between the Higgs phases and the other phases of the theory?
3. What are the observables of the Higgs phase and what is their non-perturbative behavior?

In other terms, we are inquiring what is the global *phase diagram* of a theory of gauge fields coupled to matter fields. We will see that the global properties of the phase diagram , which embody the non-perturbative behavior of the theory, depends on whether the matter carries the fundamental charge of the gauge group or not. This is the problem we will discuss now.

For simplicity (and conciseness), we will consider a theory with a gauge group $U(1)$ and a matter field that is a complex scalar field that carries charge $q \in \mathbb{Z}$. With some exceptions (that will be noted below) any of our results apply to any theory with a compact gauge group G. Also, we will not consider the case in which the matter field is fermionic. This analysis assumes that all symmetries of the theory are gauged and that there are no global symmetries (which may be unbroken or spontaneously broken.)

The action of the theory (defined on a hypercubic lattice in D Euclidean dimensions is

$$-S = \sum_{\text{links}} \beta \cos(\Delta_\mu \theta - qA_\mu) + \sum_{\text{plaquettes}} K \cos F_{\mu\nu} \tag{18.112}$$

Here the complex scalar field is $\phi = e^{i\theta}$ and the gauge field is $\mathcal{U}_\mu = e^{iA_\mu}$. The operator Δ_μ is the right lattice difference, and $F_{\mu\nu}$ is the flux of the gauge field A_μ on each plaquette. As before, K is related to the the gauge coupling constant by $K = 1/g^2$.

The phase diagram can be mapped in the $\beta - K$ plane, with $\beta > 0$ and $K > 0$. We can, and will, figure out much of the global features of the phase diagram by considering extreme regimes and then extrapolating to the middle of the diagram. However, the arguments that we present below can be shown to be rigorously correct within the radius of convergence of well defined expansions of which we will capture only their leading terms.

A. Let us consider first the regime $g^2 \to 0$ (or, equivalently, $K \to \infty$) and β finite. In this limit, the gauge fields are weakly coupled, and dominated by the flat field configurations, $F_{\mu\nu} \simeq 0$ (modulo 2π). In the extreme regime, this corresponds to the pure matter theory with a global $U(1)$ symmetry. For general dimension D, this theory has two phases, separated by a phase transition at some value β_c,

1. For $\beta < \beta_c$, the global $U(1)$ symmetry is unbroken, $\langle e^{i\theta} \rangle = 0$, and the correlation functions of the scalar field are short-ranged and decay exponentially with distance,

$$\langle e^{i\theta(x)} e^{-i\theta(0)} \rangle \sim e^{-|x|/\xi} \tag{18.113}$$

   where ξ is the correlation length. Hence, in this phase, the complex scalar field is massive and its mass is $m \sim 1/\xi$.
2. For $\beta > \beta_c$, the global $U(1)$ symmetry is spontaneously broken, $\langle e^{i\theta} \rangle \neq 0$. If the dimension $D > 2$, the correlation function has the asymptotic behavior

$$\langle e^{i\theta(x)} e^{-i\theta(0)} \rangle \sim \left| \langle e^{i\theta} \rangle \right|^2 + O(1/|x|^{D-2}) \tag{18.114}$$

   In this phase, the global $U(1)$ symmetry is spontaneously broken and the theory has a massless excitation, the Goldstone boson of the broken symmetry.

   This analysis applies to any theory with a continuous global symmetry G, provided $D > 2$. For $D = 2$, e.g. the non-linear sigma models, we already saw before that they flow to strong coupling, and in that case the global symmetry is actually unbroken, and hence, $\beta_c \to \infty$. The $U(1)$ case is special in that it has a phase transition at a finite value β_c (known as the Kosterlitz-Thouless transition), and the global $U(1)$ symmetry is also unbroken for $\beta > \beta_c$, but the correlators exhibit power-law decays: this theory has a line of fixed points. For g small but finite, we expect a Higgs phase if $\beta > \beta_c$, and a Coulomb phase (i.e. massive matter fields and a massless photon) for $\beta < \beta_c$. We will see shortly if these expectations are actually met. On the other hand, for $D > 1$, in theories with a discrete global symmetry (e.g. the Ising and $\mathbb{Z}_N$ models) the phase with have a spontaneously broken symmetry for $\beta > \beta_c$.

B. We now consider the regime $\beta \to 0$ and $K = 1/g^2$ fixed. In this regime, the matter fields are massive, with a large mass $m \approx |\ln\beta$. In this regime we expect the matter fields to decouple, and have a pure gauge theory. Again, for a continuous gauge group G, and for $D > 4$, the gauge theory has a Coulomb phase with massless gauge fields for $g < g_c$ ($K > K_c$), and a confined phase in the opposite regime, $g > g_c$. As we saw before, in the confined phase of the pure gauge theory the Wilson loop has an area law behavior, and in the deconfined phase it has a perimeter law behavior. This phase structure also holds in $D = 4$ dimensions for the $U(1)$ gauge theory. For theories with a discrete gauge group, the same phases arise for $D > 2$, except that in their deconfined phases all excitations are massive. We will see below that the deconfined phases of discrete gauge theories are topological field theories.

The results of A) and B) are summarized in the phase diagram of Fig.18.14.

C. Let us now focus on the regime in which the gauge theory is strongly coupled, $g \to \infty$ (or $K \to 0$) with β fixed. In this regime the fields fluctuate wildly. In the regime with β small we saw that, if the matter field carries the fundamental charge, $q = 1$, the expectation value of the Wilson loop (with the fundamental, and in fact, any charge) has a crossover from an area law behavior to a perimeter law for large enough loops. We saw that this crossover is compatible with theory being in a confined phase. Furthermore, in the unitary gauge, $\theta = 0$, which is always globally well defined, we see that as $g \to \infty$, the action reduces to a sum over independent link variables A_μ. The partition function of the theory is free of singularities on a vertical strip parallel to the β axis. Hence, the crossover behavior of the Wilson loop should extend along the entire strip, all the way to $\beta \to \infty$.

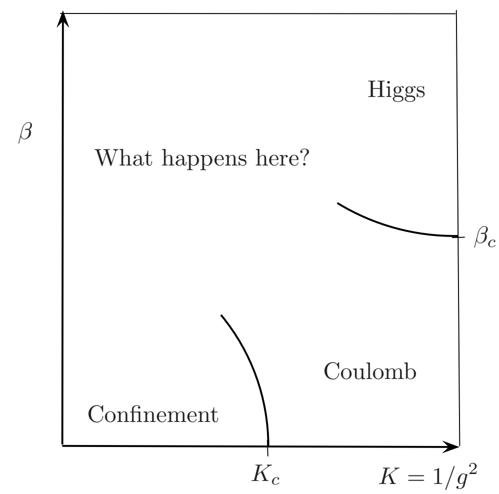

Figure 18.14 Tentative phase diagram resulting from the behaviors of A) the weakly coupled gauge theory and B) the heavy matter field.

D. Let us now inquire on the behavior of the theory for large β as a function of K. We will now see that the behavior depends on the charge q carried by the matter field. In the unitary gauge, $\theta = 0$, the action takes the simpler form

$$-S = \sum_{\text{links}} \beta \cos(qA_\mu) + \sum_{\text{plaquettes}} K \cos F_{\mu\nu} \tag{18.115}$$

The behavior for large β now depends on whether $q = 1$ or $q \neq 1$.

$q = 1$. In this case, in the limit $\beta \to \infty$ the link term of the action forces the gauge fields to be $A_\mu = 0$ (modulo 2π) on every link of the lattice. Hence, the partition function is also free of singularities on a strip along the horizontal axis at $\beta \to \infty$. This is a starling result since this "strip of analyticity" ranges from the Higgs phase at large β and small g all the way to the confining regime at large g and small β. It can be proven rigorously that this strip of analyticity exists for all theories with a compact gauge group (abelian or non-abelian, discrete or continuous) if the matter field carries the fundamental charge of the gauge group. In all such theories, there is no global distinction between a Higgs phase and a confined phase. This result is sometimes called Higgs-Confinement complementarity.

On the other hand, the Coulomb phase is also stable since in this phase there is a finite-energy state (a "particle") that carries the fundamental quantum number of the gauge group. In the case of the $U(1)$ theory, in the Coulomb phase there is a state in the spectrum that carries the $U(1)$ charge, the electron of QED. In the case of the discrete gauge groups, e.g. $\mathbb{Z}_2$, there is a finite energy state that carries the $\mathbb{Z}_2$ charge but the operator that creates this state can be shown to be non-local. We will see shortly that this is related to the fact that the deconfined phases of discrete gauge theories are topological.

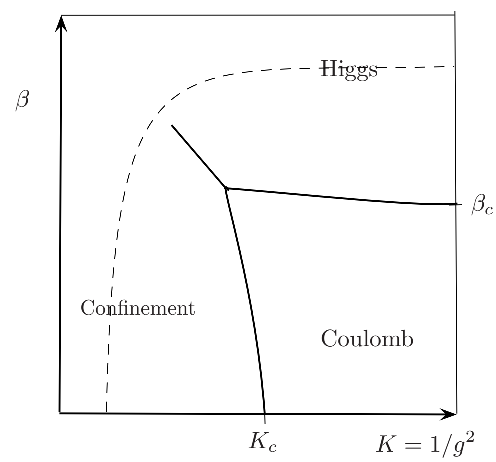

Figure 18.15 Phase diagram for a $U(1)$ gauge-matter theory with a matter theory carrying the fundamental, $q = 1$, charge of the gauge field. The broken curve represents the strip of analyticity mentioned in the text. The dot is a critical endpoint, similar to the critical point of water.

We can now summarize our understanding for the theory with a matter field that carries the fundamental gauge charge in the form of a complete phase diagram, shown in Fig.18.15. The theory has two phases: a) a Coulomb phase, and b) a Confinement-Higgs phase. In this theory confinement and Higgs are not separate phases: the theory is confining everywhere, except in the Coulomb phase. This phase diagram applies to the case of the compact $U(1)$ gauge theory coupled to a fundamental scalar (i.e. the compact abelian Higgs model) in $D = 4$ dimensions.

The nature of the phase transitions is a more subtle problem. The renormalization group analysis predicts a runaway flow into the a regime where the symmetry is broken. This predicts that the transition Coulomb-Higgs transition is weakly first order. The confinement-Coulomb transition is known to be first order from numerical simulations. The dot in Fig. 18.15 is a critical endpoint, at the end of a line of first order transitions, similar to the critical point of water. In other words, while there is no global distinction between confinement and Higgs, across the line of first order transitions local gauge-invariant observables, e.g. the flux through a local plaquette, will exhibit a jump. Interestingly, part of the phase diagram for hot and dense QCD predicts a jump from nuclear matter to the quark-gluon plasma and, hence are not separate phases.

$q > 1$. In this case, in the limit $\beta \to \infty$, the link term of the action forces the gauge field to take the one of the following possible q values on each link, $A_\mu = 2\pi n/q$, with $n = 0, 1, \ldots, q-1$. Therefore, in the limit $\beta \to \infty$, the theory reduces to a gauge theory with the discrete gauge group $\mathbb{Z}_q$. Thus, in the large β regime, the $\mathbb{Z}_q$ theory can be regarded as the low energy limit of the full theory.

If the dimension $D > 2$, the $\mathbb{Z}_q$ gauge theory has a confined phase for $K < K_c[\mathbb{Z}_q]$, and a deconfined phase in the opposite regime. Therefore, as shown in Fig.18.16, for general dimension D, the theory now has a large confined phase separate from a Coulomb phase (for β small) and a "Higgs" (deconfined) phase (for β large). Furthermore, using a Griffiths inequality, c.f. Eq.(18.72), it is straightforward to see that $K_c[\mathbb{Z}_q] \leq K_c[U(1)]$, as shown in the phase diagram of Fig. 18.16.

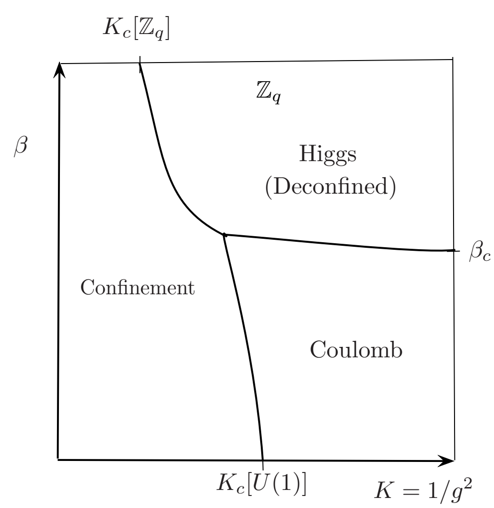

Figure 18.16 Phase diagram for a $U(1)$ gauge-matter theory with a matter theory carrying charge $q > 1$ of the gauge field. The top of the phase diagram is the $\mathbb{Z}_q$ discrete gauge theory, which has a confined and a deconfined phase. In this case there is a global distinction between the Higgs phase and the confinement phase.

The Higgs (deconfined) phase and the confined phase can be distinguished by the behavior of the Wilson loop carrying the fundamental charge, or, in fact, any charge $p < q$. As we saw above, in this case the Wilson loop with a charge $p < q$ is not algebraically screened and exhibits an area law in the confined phase. In the deconfined phase of the $\mathbb{Z}_q$ theory, for $K \gtrsim K_c[\mathbb{Z}_q]$ and large enough β, the Wilson loop exhibits a perimeter law, and the fundamental charge is indeed deconfined in this phase. This phase is a Higgs phase in the sense that the $U(1)$ gauge field is massive, although, as explained above, there isn't a local order parameter. We will see shortly that this phase is an example of a topological phase and that the low energy limit of this discrete gauge theory is a topological field theory.

This analysis also holds in the case of a non-abelian gauge theory with a matter field carrying a charge different that the fundamental charge. For example, if the gauge group is $SU(N)$ and the matter field carries the *adjoint* representation of $SU(N)$, the "unbroken" sector of the gauge group is discrete gauge group $\mathbb{Z}_N$ which is the center of the group $SU(N)$, and the adjoint representation of $SU(N)$ is blind under the action of the center, $\mathbb{Z}_N$, of the group.