א(
ו(
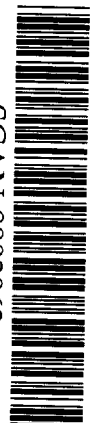
CLRC

Nonabelian Generalization of Electric-Magnetic Duality – A Brief Review

Chan Hong-Mo and Tsou Sheung Tsun

5th February 1999

Nonabelian Generalization of Electric-Magnetic Duality—a Brief Review

CHAN Hong-Mo
chanhm @ v2.rl.ac.uk
Rutherford Appleton Laboratory,
Chilton, Didcot, Oxon, OX11 0QX, United Kingdom

TSOU Sheung Tsun
tsou @ maths.ox.ac.uk
Mathematical Institute, University of Oxford,
24-29 St. Giles', Oxford, OX1 3LB, United Kingdom

Abstract

A loop space formulation of Yang-Mills theory high-lighting the significance of monopoles for the existence of gauge potentials is used to derive a generalization of electric-magnetic duality to the nonabelian theory. The result implies that the gauge symmetry is doubled from $SU(N)$ to $SU(N) \times \widetilde{SU}(N)$, while the physical degrees of freedom remain the same, so that the theory can be described in terms of either the usual Yang-Mills potential $A_\mu(x)$ or its dual $\tilde{A}_\mu(x)$. Nonabelian 'electric' charges appear as sources of A_μ but as monopoles of $\tilde{A}_\mu$, while their 'magnetic' counterparts appear as monopoles of A_μ but sources of $\tilde{A}_\mu$. Although these results have been derived only for classical fields, it is shown for the quantum theory that the Dirac phase factors (or Wilson loops) constructed out of A_μ and $\tilde{A}_\mu$ satisfy the 't Hooft commutation relations, so that his results on confinement apply. Hence one concludes, in particular, that since colour $SU(3)$ is confined then dual colour $\widetilde{SU}(3)$ is broken. Such predictions can lead to many very interesting physical consequences which are explored in a companion paper.

The question whether the electric-magnetic duality of electromagnetism is generalizable to nonabelian Yang-Mills theories is of course a classic theoretical problem of fundamental interest in its own right. Recently, however, this long-standing question has been given a new urgency by the realization that its application to the Standard Model in particle physics can lead to an understanding for the existence both of the Higgs fields required for symmetry breaking and of the three generations of fermions experimentally observed, besides offering at the same time an explanation for the values of some of the Standard Model's many empirical parameters.

In this paper we briefly review the steps leading to a solution to this problem suggested a couple of years ago. We think such a review is worthwhile since the material which has been collected over many years is scattered widely in the literature. The problem of duality in gauge theories is seen to be intimately related to the existence or otherwise of monopoles, which in turn are best described in loop space. Our present review will therefore take us over these few subjects in turn. We shall not cover, however, any of the phenomenological applications of nonabelian duality so as to avoid confusing theoretical with practical issues. Interested readers are referred to a companion paper [1] for a review of the phenomenological applications.

1   Loop Space

Most of us have learned by experience to work with gauge theory using the gauge potentials $A_\mu(x)$ as variables. But suppose we were to approach gauge theory now for the first time, we would probably ask ourselves the question whether $A_\mu(x)$ are the right variables to use to describe gauge theory. After all, $A_\mu(x)$ are gauge-dependent and therefore physically unobservable. Would it not be wiser instead to describe a physical theory with measurable quantities?

Indeed, in classical electrodynamics, the variables used by Faraday and Maxwell were not the gauge potentials $A_\mu(x)$ but the gauge invariant, measurable field strengths $F_{\mu\nu}(x)$. It was only when we started to deal with quantum mechanics that we were forced to turn to $A_\mu(x)$ as variables. That this is so is demonstrated by the famous Bohm-Aharonov experiment [2], as illustrated in Figure 1. Although the field strength $F_{\mu\nu}(x)$ vanishes throughout the region traversed by the charged particle, there are observable effects of the magnetic field H in the form of diffraction patterns on the screen, showing that $F_{\mu\nu}(x)$ itself is inadequate to describe completely the physical conditions.

The Bohm-Aharonov experiment shows that to describe the quantum mechanics of a charged particle interacting with an electromagnetic field,
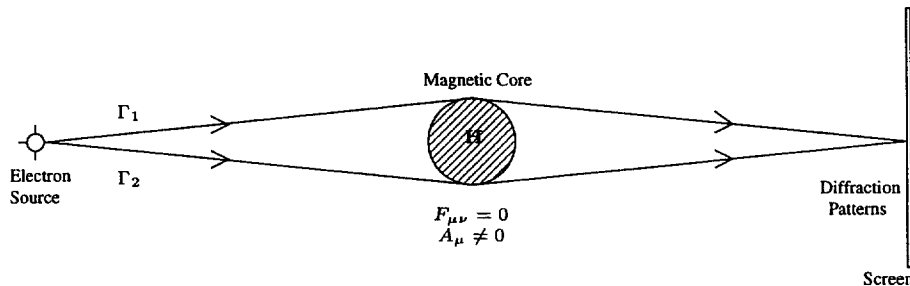
Figure 1: Schematic Representation of the Bohm-Aharonov Experiment.

the field strengths $F_{\mu\nu}(x)$ are inadequate and the gauge potentials $A_\mu(x)$ are sufficient. But the potentials $A_\mu(x)$ actually give us more information than we need. To describe the diffraction pattern on the screen, it is already sufficient to know the loop integrals:

$$\alpha_C = e \oint_C A_\mu(x)\, dx^\mu, \tag{1.1}$$

over closed paths C, not necessarily the $A_\mu(x)$'s themselves. Indeed, even α_C are more than necessary, for if all α_C change by integral multiples of 2π, the diffraction pattern will not be affected. Thus, what we need are only the phase factors:

$$\Phi_C = \exp ie \oint_C A_\mu(x)\, dx^\mu. \tag{1.2}$$

Hence, we conclude, in the words of Wu and Yang, "The field strength $F_{\mu\nu}$ underdescribes electromagnetism, ...the phase (α_C) overdescribes electromagnetism.... What provides a complete description that is neither too much nor too little is the phase factor (Φ_C)."[3] By 'underdescription' here, we mean that different physical conditions may correspond to the same values of the variables, while by 'overdescription', that different values of the variables may correspond to the same physical condition. Hence to have a unique labelling for the physical conditions in terms of $A_\mu(x)$, one will need to factor out the classes of $A_\mu(x)$ which are physically equivalent. In contrast, what is nice about the phase factors $\Phi(C)$ is that the same physical condition corresponds to the same values of $\Phi(C)$, and different conditions to different values, although any given set of values of $\Phi(C)$ need not necessarily correspond to a physical condition.

The situation in nonabelian Yang-Mills theories is similar, except that here, even in the classical theory, the field strengths $F_{\mu\nu}(x)$ no longer offer a sufficient description [4]. For the quantum theory, the gauge potentials $A_\mu(x)$ are again adequate but overdescribe the theory, and what provides a complete description for the theory yet not an over-description are the

2

path-ordered phase factors (Wilson loops):

$$\Phi_C = P \exp ig \oint_C A_\mu(x)\, dx^\mu. \tag{1.3}$$

Why then do we not use Φ_C as variables to describe gauge theory? The reason is that Φ_C is labelled by the loops C in space-time which are infinitely more numerous than the points x in space-time. Since the gauge potentials $A_\mu(x)$ labelled by x are already sufficient to describe gauge theory, a description in terms of Φ_C must therefore be highly redundant. By 'redundant' here, we mean that not all points in the space spanned by the variables $\Phi(C)$ correspond to actual physical conditions, but only a subset of it which we may think of as a constraint surface in that space. The 'redundancy' being infinite, the constraint required is bound to be complicated, which makes the description in terms of loop variables extremely clumsy. Hence, in usual circumstances, one would much rather deal with the vagaries of the gauge-dependent A_μ than with the redundancy of Φ_C. But there are situations some of which we shall discuss, where a description in terms of Φ_C is preferable, indeed may even be necessary. In that case we shall need to face the complications and develop the formalism for dealing with gauge theory in terms of loop quantities.

To effect a loop space formulation of gauge theory, our first task would be to label the loops in space-time, or in other words to introduce some sort of co-ordinates in loop space. (It will be seen that it is sufficient to consider only those loops passing through some fixed reference point $P_0 = \xi_0^\mu$.) An obvious possibility is to label a loop by the space-time co-ordinates of the points on it, thus:

$$C: \quad \{\xi^\mu(s) : s = 0 \to 2\pi, \ \xi(0) = \xi(2\pi) = \xi_0\}, \tag{1.4}$$

so that Φ_C in (1.2) or (1.3) can be rewritten as:

$$\Phi[\xi] = P_s \exp ig \int_0^{2\pi} ds\, A_\mu(\xi(s))\dot{\xi}^\mu(s), \tag{1.5}$$

where a dot denotes differentiation with respect to the parameter s. This labelling, however, is again redundant in that if one replaces s by another parametrization $s' = f(s)$, it would leave the phase factor Φ_C invariant. To effect a unique labelling of C, these reparametrizations should in principle be factored out. Nevertheless, this quotient space is so complicated that most people would rather live with the redundancy of the space of the loops parametrized by the functions ξ. This is the attitude that we shall adopt. In *parametrized loop space*, loop quantities such as $\Phi[\xi]$ are just functionals of the continuous (piece-wise smooth) functions ξ of s, which are relatively easy

to handle, although care has always to be taken in removing the additional redundancy introduced by the parametrization.

Thus, for example, a derivative can be introduced in (parametrized) loop space just as the functional derivative with respect to $\xi(s)$. To be specific, we shall define the derivative as:

$$\delta_\mu(s)\Psi[\xi] = \lim_{\Delta \to 0} \frac{1}{\Delta}\{\Psi[\xi'] - \Psi[\xi]\}, \tag{1.6}$$

with:

$$\xi'^\alpha(s') = \xi^\alpha(s') + \Delta\delta_\mu^\alpha\,\delta(s - s'), \tag{1.7}$$

meaning that the loop is 'plucked' by a delta function in the direction μ at the point on the loop corresponding to the value s of the parameter. This definition has to be interpreted with some care, especially when approximating the continuum by a discretized space as in lattice theories, where a careless handling may easily lead, for example, to asymmetric second derivatives [5, 6, 7].

Our next task in a loop space formulation is to select the variables for describing the gauge field. In doing so, we have to bear in mind a major problem already mentioned before in connection with the high degree of redundancy in loop variables. For example, suppose we choose the phase factors $\Phi[\xi]$. If we allow all of these Φ's to take any value in the gauge group G, then clearly not all of them will be expressible in terms of a local gauge potential $A_\mu(x)$ via (1.5), there not being enough freedom in $A_\mu(x)$ to satisfy all the conditions thereby imposed. In other words, there are certain sets of values of the variables $\Phi[\xi]$ in G which are unphysical. Thus, in order to ensure that in changing to a loop description we are still dealing with the same theory though in a different language, we have to impose contraints on the values that these loop variables can take so as to guarantee that one can recover a local gauge potential $A_\mu(x)$ from them. The ability to write down the appropriate constraints for doing so is thus one of the first consideration in any loop formulation of gauge theory.

For this reason, instead of the seemingly more natural choice of $\Phi[\xi]$ as variables, we choose rather to work with the quantities $F_\mu[\xi|s]$ first introduced by Polyakov [8] as the logarithmic loop derivatives of the phase factors $\Phi[\xi]$ in (1.5), namely

$$F_\mu[\xi|s] = \frac{i}{g}\Phi^{-1}[\xi]\,\delta_\mu(s)\Phi[\xi], \tag{1.8}$$

Its meaning in space-time is illustrated in Figure 2, where it can be seen that $F_\mu[\xi|s]$ depends only on that part of the loop before the point labelled by s (hence the special notation $[\xi|s]$ for its argument).

The virtue of the quantity $F_\mu[\xi|s]$ lies in the fact that in loop space it plays the role of a sort of 'connection' similar to that played by the gauge
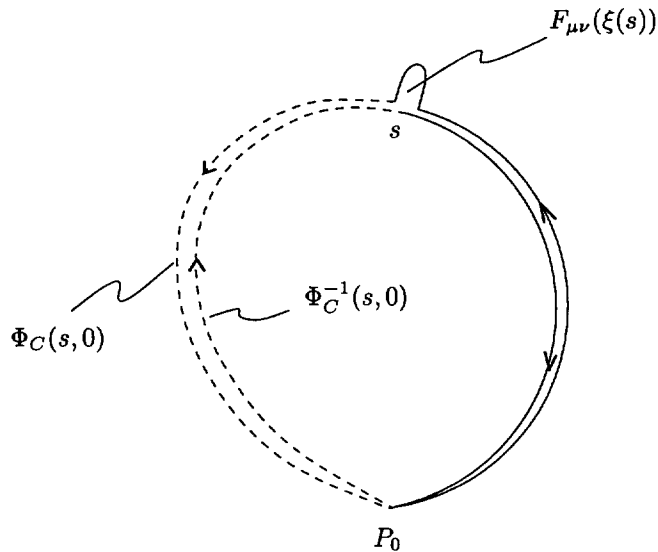
Figure 2: Illustration for the quantity $F_\mu[\xi|s]$

potential $A_\mu(x)$ in space-time. By (1.8) it tells us how the phase of $\Phi[\xi]$ changes as one moves from one loop to a neighbouring loop, i.e. from point to neighbouring point in loop space. Of course, when the phase factor $\Phi[\xi]$ exists as a single-valued function in loop space, then $F_\mu[\xi|s]$ as given in (1.8) is trivial as a 'connection', corresponding just to what is known in gauge theory language as 'pure gauge'. But for $F_\mu[\xi|s]$ as variables taking arbitrary values the corresponding connection is not in general trivial. Because of this geometrical significance, the redundancy-removing contraints take on a particularly elegant and physically lucid form, which is intimately related to the concept of monopoles as topological obstructions in gauge theories. The explicit formulation of these constraints has thus to be postponed to Section 3 after the concept of monopoles has been introduced. For the moment, we must first prepare some necessary tools.

Given the concept of $F_\mu[\xi|s]$ as a 'connection' in loop space, one can proceed as usual to define a loop space curvature as:

$$G_{\mu\nu}[\xi|s] = \delta_\nu(s)F_\mu[\xi|s] - \delta_\mu(s)F_\nu[\xi|s] + ig[F_\mu[\xi|s], F_\nu[\xi|s]]. \qquad (1.9)$$

This is the exact parallel of the familiar formula for the field strength $F_{\mu\nu}(x)$ in terms of the gauge potential $A_\mu(x)$, and has the same geometric significance of a parallel phase transport around an infinitesmal circuit, but now in loop space. Its meaning in space-time, however, is as illustrated in Figure 3 where the loop 'skips' over a small 3-volume in space. For the pure gauge connection in (1.8), the curvature $G_{\mu\nu}[\xi|s]$ is of course zero. But in general the curvature need not vanish, in which case, as will be shown in the next
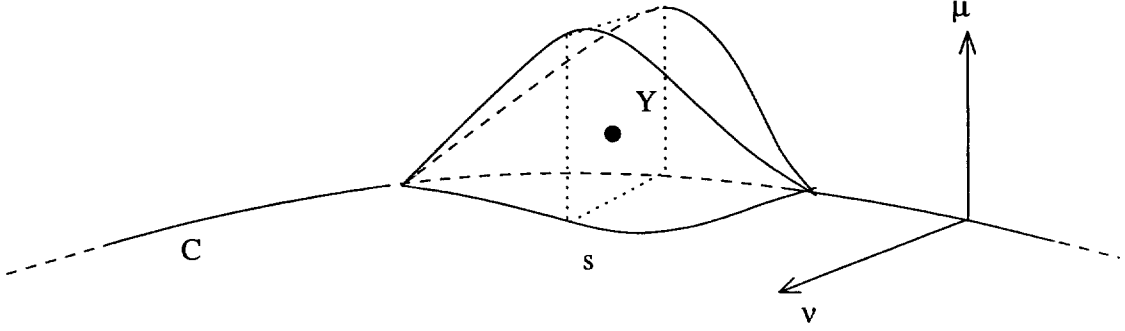
5

Figure 3: Illustration for $G_{\mu\nu}[\xi|s]$.

section, it means physically that there is a monopole charge enclosed inside the small 3-volume 'skipped' over by the loop in Figure 3.

Further, just as one has constructed from the potential $A_\mu(x)$ the phase factor $\Phi[\xi]$ which, as the 'holonomy', is an extension of the concept of curvature to a finite-sized loop, so a holonomy in loop space can also be constructed from the 'connection' $F_\mu[\xi|s]$ as [9]:

$$\Theta_\Sigma = P_t \exp ig \int_0^{2\pi} dt \int_0^{2\pi} ds \, F_\mu[\xi_t|s] \frac{\partial \xi_t^\mu(s)}{\partial t}, \qquad (1.10)$$

where Σ denotes the parametrized surface:

$$\Sigma : \{\xi_t^\mu(s) \colon s = 0 \to 2\pi, t = 0 \to 2\pi\}, \qquad (1.11)$$

with:

$$\xi_t^\mu(0) = \xi_t^\mu(2\pi) = \xi_0^\mu, \ t = 0 \to 2\pi, \qquad (1.12)$$

$$\xi_0^\mu(s) = \xi_{2\pi}^\mu(s) = \xi_0^\mu, \ s = 0 \to 2\pi. \qquad (1.13)$$

The closed surface Σ swept out by the one-parameter family of loops ξ_t is illustrated in Figure 4, which may be considered also as a loop in loop space. Again, for the 'pure gauge' connection (1.8), Θ_Σ is trivial and equals the group identity. However, it will not be so when the volume enclosed by Σ contains monopole charges, as we shall see later.

To round off this section, we mention some facts which we shall find useful later. From Figure 2, it can be seen that in terms of ordinary field variables, $F_\mu[\xi|s]$ is also expressible as [8, 10]:

$$F_\mu[\xi|s] = \Phi_\xi^{-1}(s,0) F_{\mu\nu}(\xi(s)) \Phi_\xi(s,0) \dot{\xi}^\nu(s), \qquad (1.14)$$

where $\Phi_\xi(s_2, s_1)$ is the parallel phase transport:

$$\Phi_\xi(s_2, s_1) = P_s \exp ig \int_{s_1}^{s_2} ds \, A_\mu(\xi(s)) \dot{\xi}^\mu(s) \qquad (1.15)$$

6
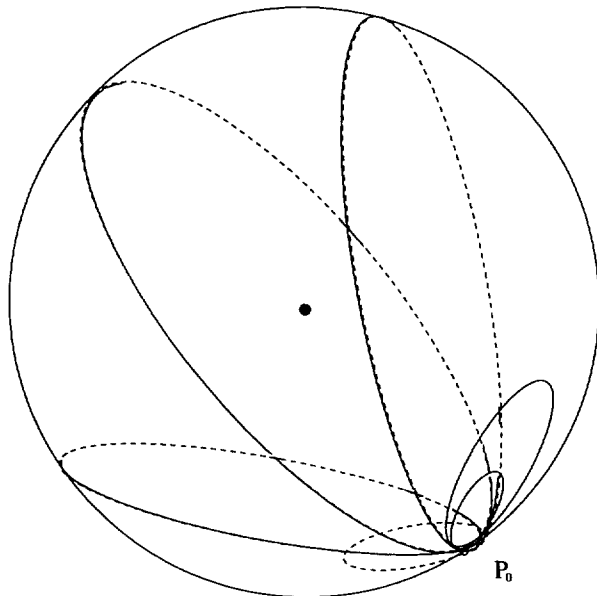
Figure 4: Surface swept out by the one-parameter family of loops ξ_t.

from s_1 to s_2 along the loop ξ. This formula (1.14) allows us to translate from loop space language back to local field language. For example, by substitution of (1.14) into the expression:

$$\mathcal{A}_F^0 = -\frac{1}{4\pi\bar{N}} \int \delta\xi \int_0^{2\pi} ds \operatorname{Tr}\{F_\mu[\xi|s]F^\mu[\xi|s]\}|\dot{\xi}(s)|^{-2}, \qquad (1.16)$$

and performing the functional integral, one obtains the standard pure Yang-Mills action:

$$-\frac{1}{16\pi} \int d^4x \operatorname{Tr}\{F_{\mu\nu}(x)F^{\mu\nu}(x)\} \qquad (1.17)$$

in terms of local field variables, if we define the normalization factor $\bar{N}$ as:

$$\bar{N} = \int_0^{2\pi} ds \int \prod_{s'\neq s} d^4\xi(s'). \qquad (1.18)$$

The expression (1.16) will serve as the loop space field action in terms of $F_\mu[\xi|s]$ as variables.

2 Monopoles

Monopoles occur as topological obstructions in gauge theories with compact multiply-connected gauge groups. They may be defined as follows [11, 3, 12].

Take a 1-parameter family of closed loops $\{C_t\}$, as that parametrized by ξ_t in (1.11) above, sweeping out a 2-dimensional surface Σ. For each t we can then associate a phase factor $\Phi(C_t)$ which is an element of the gauge group G. As t varies from 0 to 2π, $\Phi(C_t)$ traces out a closed curve, say Γ_Σ, in G. If G is multiply-connected, then Γ_Σ will belong to one or other of the homotopy classes $\pi_0(G)$ of closed curves in G, where members of different classes cannot be continuously deformed into one another. The homotopy class to which Γ_Σ belongs is defined as the monopole charge enclosed inside the surface Σ.

At first sight, this might seem a rather abstruse definition for a monopole for which, after all, the primary example is just the magnetic charge of electromagnetism, which can be represented simply by a novanishing divergence of the magnetic field, or in relativistic notation by a nonvanishing $\partial_\mu{}^*F^{\mu\nu}(x)$. On closer examination, however, it is easily seen first, that the above definition reduces in the abelian theory back to the usual interpretation of the monopole as a source of the field $^*F$, and secondly, that for a nonabelian theory this latter interpretation no longer works and cannot be used to define the monopole [7]. Indeed, the above definition, or its equivalent, is the only known valid extension of Dirac's magnetic monopole to nonabelian Yang-Mills theory.

This definition exhibits the essentially topological nature of the monopole charge which is by definition discrete, and since invariant under continuous deformations, also conserved. The values that this charge can take depend on the topological property of the gauge group. Thus, for (compact) electrodynamics, the gauge group is $U(1)$ which has the topology of a circle, on which the homotopy classes of closed curves are labelled by their winding numbers. As a result, the magnetic charge is quantized, meaning that it takes integral values in $\mathbf{Z}$, as first noted by Dirac [13].

For simply-connected gauge groups such as $SU(N)$, there can be no monopoles, there being only one homotopy class of closed curves which contains the vacuum. However, this does not mean that there can be no nonabelian monopoles in gauge theories with $su(N)$ symmetries. By an $su(N)$ theory, one usually means a theory invariant under the gauge Lie algebra $su(N)$. This by itself does not specify the gauge group, since different Lie groups can correspond to the same Lie algebra. But it is the the global structure of the gauge group which determines whether a theory can have monopoles. Thus, for the pure $su(N)$ Yang-Mills theory containing only gauge bosons in the adjoint representation and nothing else, the gauge group is $SU(N)/\mathbf{Z}_N$ and not $SU(N)$, since two elements in $SU(N)$ differing by only a factor $\exp 2i\pi/N$ have the same effect on the gauge boson field and should thus be considered as identical elements of the gauge group. That being the case, and $SU(N)/\mathbf{Z}_N$ being N-tuply connected, the pure $su(N)$ Yang-Mills theory can

have monopoles with charges labelled by elements of $\mathbf{Z}_N$. In particular, pure $su(2)$ Yang-Mills theory has gauge group $SU(2)/\mathbf{Z}_2 = SO(3)$ and monopole charges labelled by a sign $\pm$, with $+$ corresponding to the vacuum and the charge $-$ being its own conjugate. Similarly, pure $su(3)$ Yang-Mills theory has gauge group $SU(3)/\mathbf{Z}_3$ and monopole charges labelled by the cube roots of unity $1, \omega, \omega^2$, with $\omega = \exp 2\pi i/3$.

To determine the gauge group and hence whether a theory has monopoles, we need to examine the gauge transformation properties of all fields present in the theory [14, 15]. Take for example the electroweak theory as we know it today which has either:

(i) $SU(2)$ doublets with half-integral hypercharges, e.g. $(\nu, e)_L$ with $Y = 1/2$; or else:

(ii) $SU(2)$ singlets or triplets with integral hypercharges, e.g. e_R with $Y = 1$, A_μ with $Y = 0$.

Hence, if we put:

$$\tilde{f} = \exp 2\pi i T_3 \quad \in SU(2)_f, \tag{2.1}$$

and:

$$\tilde{y} = \exp 2\pi i Y \quad \in U(1)_Y, \tag{2.2}$$

the couple $(f\tilde{f}, y\tilde{y})$ in $SU(2)_f \times U(1)_Y$ has exactly the same physical effect as (f, y) and therefore has to be identified with the latter. As a result, the gauge group is not $SU(2) \times U(1)$ but $[SU(2) \times U(1)]/\mathbf{Z}_2 = U(2)$. Now, in contrast to $SU(2)$, the group $U(2)$ can have monpoles [15]. Indeed, as seen in Figure 5, AB is a closed curve in $U(2)$ which cannot be continuously deformed to zero. It winds half-way round each of the $SU(2)_f$ and $U(1)_Y$ subgroups. Hence a $U(2)$-monopole of unit charge can be thought of as carrying an $SO(3)$ monopole charge $\eta = -$, as well as a $U(1)_Y$ monopole charge of half the Dirac value. In general, the monopoles of the electroweak theory are labelled by an integer n, where a charge n monopole can be thought of as carrying simultaneously:

$$\eta = (-1)^n \quad SO(3) \text{ monopole charge,}$$
$$\tilde{y} = \pi n/g_1 \quad U(1)_Y \text{ monopole charge.} \tag{2.3}$$

A similar analysis carried out for the full Standard Model given the presently known spectrum of charges reveals that the gauge group is $[SU(3) \times SU(2) \times U(1)]/\mathbf{Z}_6$, and that its monopoles, also labelled by an integer n, can be thought of as carrying simultaneously the following charges [15]:

$$\zeta = \exp 2\pi i n/3 \quad SU(3)/\mathbf{Z}_3 \text{ monopole charge,}$$
$$\eta = (-1)^n \quad SO(3) \text{ monopole charge,}$$
$$\tilde{y} = 2\pi n/(3g_1) \quad U(1)_Y \text{ monopole charge.} \tag{2.4}$$

9
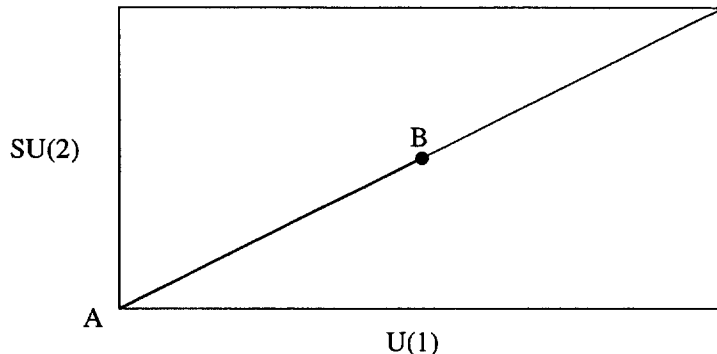
Figure 5: Monopoles in the electroweak theory

This fact will be of use in the physical applications of nonabelian duality [1].

The lists given in (2.3) and (2.4) of available monopole charges in respectively the electroweak theory and the Standard Model are the equivalents of the statement in electromagnetism that the magnetic charge is quantized. The well-known Dirac quantization condition for the abelian theory, however, contains more information, for it says not only that the magnetic charge $\tilde{e}$ is quantized but that it is quantized in units of $1/2e$, which can be thought of as a relation

$$2e\tilde{e} = 1 \qquad (2.5)$$

between the minimal coupling strengths. A parallel for this exists also for $su(N)$ theories which for our present normalization convention reads as [25]:

$$g\tilde{g} = 1, \qquad (2.6)$$

and can be similarly derived.[1]

The definition given above for the monopole charge is unfortunately a little abstract. In order to be useful, the monopole charge has to be expressed explicitly in terms of whatever field variables one may choose to adopt. In terms of the standard variables $A_\mu(x)$, monopole charges are always a little hard to handle. This can be seen already in the abelian theory. If $A_\mu(x)$ exists and is single-valued, then it follows that:

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0, \qquad (2.7)$$

or that for:

$$^*F_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}, \qquad (2.8)$$

---

[1]The couplings here which follow the original Dirac convention are the so-called unrationalized couplings. For the rationalized couplings now more in common use, the conditions should read respectively $e\tilde{e} = 2\pi$ and $g\tilde{g} = 4\pi$.

we have:

$$\partial_\mu{}^*F^{\mu\nu} = 0. \tag{2.9}$$

However, in the presence of a monopole, $\partial_\mu{}^*F^{\mu\nu}$ cannot vanish. Hence, $A_\mu(x)$ must be singular somewhere. This is the reason for the Dirac string [13].

The way out, of course, is to consider $A_\mu(x)$ as a patched quantity [3]. One covers, for example, the sphere by two patches (N and S):

$$
\begin{aligned}
(N): \quad & 0 \leq \theta < \pi, \quad & 0 \leq \phi < 2\pi, \\
(S): \quad & 0 < \theta \leq \pi, \quad & 0 \leq \phi < 2\pi,
\end{aligned}
\tag{2.10}
$$

and define $A_\mu^{(N)}$ and $A_\mu^{(S)}$ separately in N and S, with the two potentials related to each other in the overlap region by a gauge transformation parametrized by the (patching) function:

$$S = \exp ie\alpha, \tag{2.11}$$

where for S to be well-defined the phase α has to change by an integral multiple of 2π for $\phi = 0 \to 2\pi$, giving thus the Dirac quantization condition (2.5). Similarly for nonabelian Yang-Mills theory, one can introduce patched potentials to accommodate monopoles, the procedure being then a little more complicated. In either case, however, the patching depends on the locations of the monopoles, and the number of patches required increases exponentially with the number of monopoles introduced. It thus appears that a description of monopoles in terms of $A_\mu(x)$ is going to be very complicated, so much so that even the intrinsic difficulty of loop space formulations may now be worth facing by comparison.

The definition above of the monopole charge being given in terms of loop quantities in the first place, it is not surprising that it has a simpler representation in the loop space formulation, especially for the nonabelian theory [9, 7]. For illustration, it is sufficient to exhibit this explicitly only for the simplest example with gauge group $SO(3)$. Recall that $F_\mu[\xi|s]$ is by definition the logarithmic derivative of $\Phi[\xi]$. Hence, we may write:

$$\exp igdt \int_0^{2\pi} ds\, F_\mu[\xi_t|s](\partial\xi_t^\mu(s)/\partial t) \sim \Phi^{-1}[\xi_{t+dt}]\Phi[\xi_t]. \tag{2.12}$$

The loop space holonomy Θ_Σ is the product ordered in t of such factors and is thus the total change in $\Phi[\xi_t]$ as $t = 0 \to 2\pi$. Now both Φ and Θ may be interpreted as an element of either the gauge group $SO(3)$ or its double cover $SU(2)$. However, if we wish to exhibit the monopole charge as an element of $\mathbf{Z}_2$ considered as a subgroup of $SU(2)$, then we should work in the latter. Remembering that the corresponding curve Γ_Σ is a closed curve in $SO(3)$, we see that Θ_Σ must wind around $SU(2)$ an odd number of 'half-times' if

11
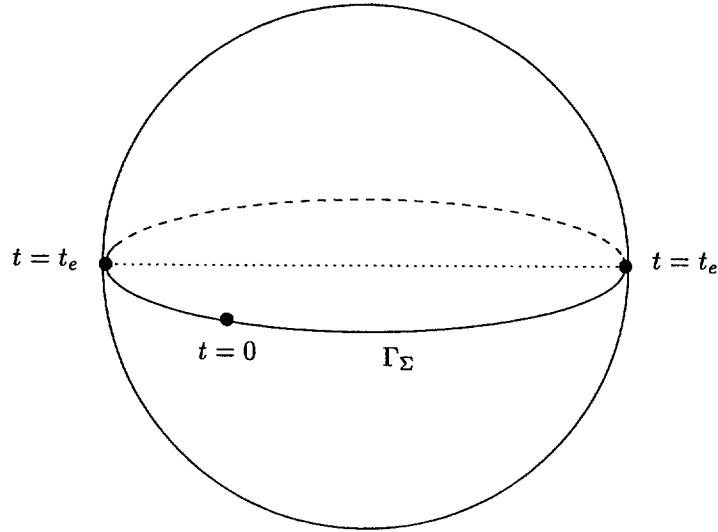
Figure 6: A curve representing an $SO(3)$ monopole

Σ contains a monopole charge $-$, but an even number of 'half-times' if Σ contains no monopole. Hence we conclude:

$$\Theta_\Sigma = \zeta_\Sigma I, \tag{2.13}$$

where ζ_Σ is the monopole charge enclosed inside Σ. This formula (2.13) actually holds for any theory with gauge group $SU(N)/\mathbf{Z}_N$.

It is instructive to examine how this result arises in detail in terms of the (patched) gauge potential (Figure 6). Without loss of generality, we shall choose the reference point $P_0 = \xi_0^\mu$ to be in the overlap region, say on the equator which corresponds to the loop ξ_{t_e}. Starting at $t = 0$, where $\Phi^{(N)}[\xi_0]$ is the identity, the phase factor $\Phi^{(N)}[\xi_t]$ traces out a continuous curve in $SU(2)$ (which has the topology of a 3-sphere) until it reaches $t = t_e$. At $t = t_e$ one makes a patching transformation and goes over to $\Phi^{(S)}[\xi_t]$. From $t = t_e$ onwards, the phase factor $\Phi^{(S)}[\xi_t]$ again traces out a continuous curve until t reaches 2π, where it becomes again the identity and joins up with $\Phi^{(N)}[\xi_0]$. In order that the curve Γ_Σ so traced out winds only half-way round $SU(2)$ while being a closed curved in $SO(3)$, as it should if Σ contains a monopole, we must have

$$\Phi^{(N)}[\xi_{t_e}] = -\Phi^{(S)}[\xi_{t_e}], \tag{2.14}$$

which means for the holonomy

$$\Theta_\Sigma = (\Phi^{(S)}[\xi_{t_e}])^{-1}\,\Phi^{(N)}[\xi_{t_e}] = -I \tag{2.15}$$

as required in (2.13).

The formula (2.13) for the monopole charge in terms of the loop space holonomy Θ, though already explicit, is not as convenient for our discussion as the differential formula in terms of the loop space curvature $G_{\mu\nu}$. As noted already in Section 1, curvature is just an infinitesimal version of the holonomy. Hence, in the absence of monopoles G vanishes, but if the loop ξ passes through a monopole at the point labelled by s, then $G_{\mu\nu}[\xi|s]$ must take on a value equal to the logarithm of the monopole charge ζ at that point. Hence, we can write, for a classical monopole moving along the world line $Y^\mu(\tau)$ [16]:

$$G_{\mu\nu}[\xi|s] = -4\pi\tilde{g} \int d\tau \kappa[\xi|s]\epsilon_{\mu\nu\rho\sigma}\dot{\xi}^\rho(s)\frac{dY^\sigma(\tau)}{d\tau}\delta^4(\xi(s) - Y(\tau)), \qquad (2.16)$$

where $\kappa[\xi|s]$ satisfies:

$$\exp i\pi\kappa = \zeta. \qquad (2.17)$$

This formula has to be interpreted with some care since given ζ the solution for κ in (2.17) in the Lie algebra is not unique. But in any case, for $\zeta \neq 1$, $\kappa \neq 0$ which means that monopoles can be regarded as sources of curvature in loop space, as already anticipated. For the abelian theory, the equation (2.16) reduces to the familiar formula for a classical point magnetic charge:

$$\partial_\nu {}^*F^{\mu\nu}(x) = -4\pi\tilde{e} \int d\tau \frac{dY^\mu(\tau)}{d\tau}\delta(x - Y(\tau)). \qquad (2.18)$$

3   Wu-Yang Criterion and Poincaré Lemma

An attractive feature of monopoles considered as topological obstructions in gauge fields is that their topology defines their own dynamics. This was first pointed out in a beautiful paper by Wu and Yang in 1976 [17]. That this is so is intuitively clear. The assertion that there is a monopole at a certain point x in space-time, as discussed in the last section, means that the gauge field surrounding x has to have a certain topological structure, and if the monopole is displaced to another point, then the gauge field will have to rearrange itself so as to maintain the same topological structure around the new point. There is thus naturally a coupling between the gauge field and the position of the monopole, or in physical language a topologically induced interaction between the field and the monopole.

To deduce the explicit form of this interaction, one can proceed as follows. One writes down first the free action of the gauge field together with that of a particle. For example, for electromagnetism and a classical particle of mass m, one has:

$$\mathcal{A}^0 = \mathcal{A}_F^0 + \mathcal{A}_M^0, \qquad (3.1)$$

13

where $\mathcal{A}_F^0$ is the usual Maxwell action, and

$$\mathcal{A}_M^0 = -\int d\tau, \tag{3.2}$$

the integral being taken along the world-line of the particle with τ being the proper time along the world-line. Extremizing this action with respect to the dynamical variables of the problem, namely $A_\mu(x)$ for the field and co-ordinates $Y^\mu(\tau)$ for the particle, one obtains the free equations of the field and of the particle. Suppose now, however, one stipulates that the particle carries a (magnetic) monopole charge and imposes on to the system the appropriate topological condition that this should be so. This condition couples the field and the particle, as already explained, so that if one extremizes again the action (3.2) under the imposed constraint, the equations of motion will no longer be free but coupled equations involving an 'interaction' between the particle and the field.

What are the coupled equations so obtained? Knowing as one does that classical electromagnetism is dual symmetric, one is not surprised that they turn out to be just the dual to the Maxwell and Lorentz equations for the motion of an electric charge in an electromagnetic field. Indeed, this was the way that Wu and Yang deduced the equations in their original paper [17]. However, a direct attack on the problem so posed is not as easy as it might seem at first sight. The reason is that $A_\mu(x)$, which is the dynamical variable for the field, is a patched quantity in the presence of a monopole, as explained in the last section, where the patching depends on the other dynamical variable $Y^\mu(\tau)$ for the particle. Extremizing $\mathcal{A}^0$ with respect to $A_\mu(x)$ is thus not a simple matter, although (according to Wu in private communication) possible.

There is, however, a very simple and elegant method for solving this problem [16, 7]. The trick is to adopt as field variables not the gauge potential $A_\mu(x)$ but the field strength $F_{\mu\nu}(x)$. Being gauge invariant, $F_{\mu\nu}(x)$ is not patch-dependent even in the presence of a monopole. Further, the topological condition defining a (magnetic) monopole charge at $Y^\mu(\tau)$ can be expressed simply in terms of $F_{\mu\nu}(x)$ as (2.18). Incorporating (2.18) as a constraint on the action (3.1) by means of a Lagrange multiplier $\lambda_\mu(x)$ and extremizing with respect to $F_{\mu\nu}(x)$ and $Y^\mu(\tau)$, one obtains then easily the equations:

$$F^{\mu\nu}(x) = 4\pi[\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}(\partial_\sigma\lambda_\rho(x) - \partial_\rho\lambda_\sigma(x))], \tag{3.3}$$

and

$$m\frac{d^2Y^\mu(\tau)}{d\tau^2} = -4\pi\tilde{e}[\partial_\nu\lambda_\mu(Y(\tau)) - \partial_\mu\lambda_\nu(Y(\tau))]\frac{dY^\nu(\tau)}{d\tau}. \tag{3.4}$$

14

The equation (3.3) says that the dual Maxwell field ${}^*F_{\mu\nu}(x)$ is also a gauge field:

$$
{}^*F_{\mu\nu}(x) = \partial_\nu \tilde{A}_\mu(x) - \partial_\mu \tilde{A}_\nu(x), \tag{3.5}
$$

with the (dual) potential:

$$
\tilde{A}_\mu(x) = 4\pi \lambda_\mu(x). \tag{3.6}
$$

Then using this, one can rewrite the other equation (3.4) as:

$$
m\frac{d^2 Y^\mu(\tau)}{d\tau^2} = -\tilde{e}\,{}^*F^{\mu\nu}(Y(\tau))\frac{dY_\nu(\tau)}{d\tau}. \tag{3.7}
$$

As expected, these equations, together with the constraint (2.18), are exactly the dual of the equations of motion for an electric charge in an electromagnetic field.

There is an important detail in the above derivation which at first sight looks like a flaw but which when understood has far reaching consequences for the development which follows. The variables $F_{\mu\nu}(x)$ adopted to solve the variational problem are more numerous than the original variables $A_\mu(x)$ (there being 6 components to $F_{\mu\nu}$ compared with 4 to A_μ) and must therefore be regarded as 'redundant' in the sense the term was used in Section 1 while discussing loop variables. In other words, given a set of values for $F_{\mu\nu}(x)$, there is no guarantee that they can be derived from a potential $A_\mu(x)$ unless their values are appropriately constrained. Now, the beauty of the above derivation is that the dynamical constraint (2.18) imposed, representing the topological definition of the monopole charge, already ensures the existence of the potential, thereby removing automatically the redundancy in the new variables. Indeed, one notices from (2.18) that except on the world-line $Y(\tau)$ of the monopole, one has:

$$
\partial_\nu {}^*F^{\mu\nu}(x) = 0, \tag{3.8}
$$

which is exactly the condition which implies that $F_{\mu\nu}(x)$ is derivable from a potential. This well-known mathematical fact, to which we shall have ample occasions to recall, we shall refer to as the Poincaré Lemma, although it is only a very special case of that important theorem in differential geometry. The only place where the condition (2.18) does not guarantee the existence of the potential is on the monopole world-line, but that is no problem since at the monopole position $A_\mu(x)$ should not exist in any case.

The same derivation can be applied also to a quantum particle carrying a monopole charge. For example, for a Dirac particle, we replace the action (3.2) above by [18]:

$$
\mathcal{A}_M^0 = \int d^4 x\, \bar{\tilde{\psi}}(x) \left(i\partial_\mu \gamma^\mu - m\right) \tilde{\psi}(x), \tag{3.9}
$$

15

and the current on the right-hand side of (2.18) by its quantum equivalent:

$$\partial_\nu \,^*F^{\mu\nu}(x) = -4\pi \tilde{e} \bar{\tilde{\psi}}(x) \gamma^\mu \tilde{\psi}(x). \tag{3.10}$$

Extremizing then the action $\mathcal{A}^0$ with respect to $F_{\mu\nu}(x)$ and $\tilde{\psi}(x)$ under the constraint (3.10) yields the equation (3.3) together with the equation [18]:

$$(i\partial_\mu \gamma^\mu - m)\tilde{\psi}(x) = -\tilde{e}\tilde{A}_\mu(x)\gamma^\mu \tilde{\psi}(x). \tag{3.11}$$

Again, the equations obtained are exactly the dual for that of an electric charge as expected.

Our next objective is now to generalize to nonabelian gauge theory to derive the equations of motion for monopoles. This is not possible using again as variables the field strength $F_{\mu\nu}(x)$ since, in contrast to the abelian theory, these are not gauge-invariant and has therefore also to be patched in the presence of a monopole. This difference with the abelian theory is of course very deep, and no simple modifications are likely to overcome this difficulty. For this reason, we turn to the loop formulation treated in Section 1 where the variables are constructed to be gauge invariant.[2] We write then the free field action $\mathcal{A}_F^0$ in terms of the loop variables $F_\mu[\xi|s]$ in the form given in (1.16), and, by the Wu-Yang criterion, impose as a dynamical constraint the condition that the particle in $\mathcal{A}_M^0$ should carry a monopole charge. Now, according to (2.16), this condition can be written in terms of the loop curvature $G_{\mu\nu}[\xi|s]$ as:

$$G_{\mu\nu}[\xi|s] = -4\pi J_{\mu\nu}[\xi|s], \tag{3.12}$$

where $J_{\mu\nu}[\xi|s]$ represents the monopole current. For a classical particle, this takes the form [16]:

$$J_{\mu\nu}[\xi|s] = \tilde{g} \int d\tau \kappa[\xi|s] \epsilon_{\mu\nu\rho\sigma} \dot{\xi}^\rho(s) \frac{dY^\sigma(\tau)}{d\tau} \delta^4(\xi(s) - Y(\tau)), \tag{3.13}$$

and for a Dirac particle, it takes the form [18]:

$$J_{\mu\nu}[\xi|s] = \tilde{g}\epsilon_{\mu\nu\rho\sigma}[\bar{\tilde{\psi}}(\xi(s))\Omega_\xi(s,0)\gamma^\rho \tau_i \Omega_\xi^{-1}(s,0)\tilde{\psi}(\xi(s))]\tau^i \dot{\xi}^\sigma(s). \tag{3.14}$$

We shall return later to explain the meaning of the operator appearing in (3.14):

$$\Omega_\xi(s,0) = \omega(\xi(s))\Phi_\xi(s,0). \tag{3.15}$$

---

[2] Actually, as defined, the variables $\Phi[\xi]$ and $F_\mu[\xi|s]$ depend on the gauge transformation at the reference point $P_0 = \xi_0^\mu$, but such a transformation is harmless as far as patching is concerned.

which will be assigned a rather significant role in the applications of nonabelian duality to phenomenology. For the moment, it suffices only to note that it represents a frame rotation in internal symmetry space.

As it stands, the variational problem posed is straightforward though somewhat complicated. Thus, incorporating the constraint (3.10) into the action by means of the Lagrange multipliers $L_{\mu\nu}[\xi|s]$:

$$\mathcal{A}' = \mathcal{A}^0 \int \delta\xi ds \, \mathrm{Tr}\{L^{\mu\nu}[\xi|s]\{G_{\mu\nu}[\xi|s] + 4\pi J_{\mu\nu}[\xi|s]\}\}, \qquad (3.16)$$

and extremizing with respect to the variables $F_{\mu\nu}[\xi|s]$ and $Y^\mu(\tau)$ or $\tilde{\psi}(x)$, one obtains the equations of motion for the nonabelian monopole. We give here only the equations for the more interesting Dirac particle [18], namely:

$$\delta_\mu(s)F^\mu[\xi|s] = 0, \qquad (3.17)$$

which, according to Polyakov [8], is the loop equivalent of the Yang-Mills field equation, and:

$$(i\partial_\mu\gamma^\mu - m)\tilde{\psi}(x) = -\tilde{g}\tilde{A}_\mu(x)\gamma^\mu\tilde{\psi}(x), \qquad (3.18)$$

where the quantity $\tilde{A}_\mu(x)$ which couples to the Dirac particle like a dual potential can again be expressed in terms of the Lagrange multiplier $L_{\mu\nu}[\xi|s]$ but now as a rather complicated functional integral.

The equations derived for nonabelian monopoles from the Wu-Yang criterion as outlined above are new since we know as yet of no nonabelian generalization to the abelian electric-magnetic duality by means of which we were able to infer in the abelian case the equations for the magnetic charge from those of the electric charge. However, we shall not examine the details of these new equations for the present, for we shall be able later to achieve a much better appreciation of them. What is of greater interest at the moment is a missing link in the derivation akin to that already encountered in the parallel derivation of the abelian equations, namely the question of the redundancy of the loop variables $F_\mu[\xi|s]$. What are the necessary constraints on $F_\mu[\xi|s]$ to ensure that one can recover from them the original $A_\mu(x)$, and have these constraints been satsified in the above treatment?

To answer this question, one has to find a generalization to the nonabelian theory of the Poincaré Lemma which was used to answer a similar question in the abelian theory. Given the large number of loop variables, this at first sight looks very difficult. Fortunately, one is able to guess an answer by following a physical intuition based on what one has learned in the abelian case [9]. We recall that by the Poincaré Lemma, what guaranteed the recovery of the potential at the point x from $F_{\mu\nu}(x)$ was the condition (3.8), which means physically that at x there is no magnetic charge. Could it not be then that

even for the nonabelian theory, the absence of monopole charge at any point x would guarantee the local existence of the gauge potential? If that is true, then our derivation is complete, for the dynamical constraint imposed (2.16) does imply that at all points outside the world-line of the monopole, the loop curvature $G_{\mu\nu}$ representing the monopole charge vanishes.

This conjecture is found to be correct. One is indeed able to derive an Extended Poincaré Lemma [9], which states that given $F_\mu[\xi|s]$ depending on ξ only up to the point s as the notation $[\xi|s]$ in the argument implies, and satisfying the transversality condition:

$$F_\mu[\xi|s]\dot{\xi}^\mu(s) = 0, \tag{3.19}$$

then a gauge potential $A_\mu(x)$ can be locally constructed in regions of space-time where the loop space curvature $G_{\mu\nu}[\xi|s]$ vanishes. That there is a transversality condition (3.19) on $F_\mu[\xi|s]$, is not surprising. As seen in (1.8), the longitudinal component of $F_\mu[\xi|s]$ represents the logarithmic variation of $\Phi[\xi]$ along the loop, just as that induced by a reparametrization of the loop. The condition (3.19) thus corresponds to the parametrization independence of the phase factor $\Phi[\xi]$ and is the price one has to pay for working in *parametrized loop space* for removing the redundancy thus introduced, as was explained in Section 1. In terms of local field variables, as seen in (1.14), it corresponds to the statement that $F_{\mu\nu}(x)$ is antisymmetric in its indices μ and ν. This additional constraint is relatively easy to take care of, affecting little the following arguments, and for this reason will largely be ignored.

Apart then from a 'proof' of the Extended Poincaré Lemma [9, 7], which we briefly outline below, the loop space formulation of nonabelian Yang-Mills theory introduced in Section 1, and the extension of the Wu-Yang criterion to the nonabelian Yang-Mills case, are now both complete.

We shall indicate how one can arrive at an Extended Poincaré Lemma only for the case when there are no monopole charges anywhere in space. This will be sufficient to illustrate the idea. For the more general case with a number of isolated monopole charges, which is of interest to the above application of the Wu-Yang criterion, some more technical sophistication is required [9] but the idea remains similar.

First, we note that if all $G_{\mu\nu}[\xi|s]$ vanish, then Θ_Σ as defined in (1.10) equals the identity for all parametrized surfaces Σ. Recalling that Σ is a loop in loop space, this implies that the following integral over an open path in loop space:

$$\Theta_\Sigma(t,0) = P_{t'}\exp ig\int_0^t dt'\int_0^{2\pi} ds F_\mu[\xi_{t'}|s]\frac{\partial\xi_{t'}^\mu(s)}{\partial t'}, \tag{3.20}$$

is path-independent and is a function only of the 'end-point' ξ_t. We define then the phase factor as:

$$\Phi[\xi_t] = \Theta_\Sigma^{-1}(t,0), \tag{3.21}$$

18
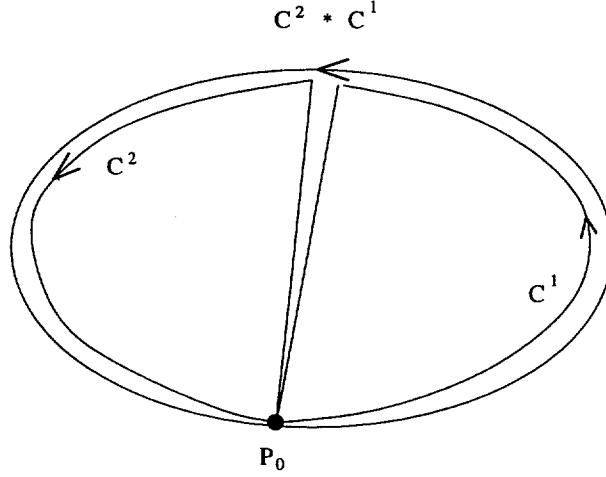
Figure 7: Illustration for the composition law for loops

which by definition gives $F_\mu[\xi|s]$ as its logarithmic loop derivative. Second, by construction, one shows, from the fact that $F_\mu[\xi|s]$ is transverse and depends on ξ only up to the point s, that $\Phi[\xi]$ obeys the composition law, namely that:

$$\Phi(C^2 * C^1) = \Phi(C^2)\Phi(C^1), \tag{3.22}$$

where C denotes the geometric loop in space-time corresponding to the parametrized loop ξ, and $C^2 * C^1$ represents the loop obtained by first going around C^1 then C^2, as illustrated in Figure 7. Third, to every point x in space-time, draw a straight line γ_x joining the reference point $P_0 = \xi_0^\mu$ to x and construct the phase factor Φ for the triangle formed by $\gamma_{x'}^{-1} * \gamma_{x'x} * \gamma_x$:

$$h(x', x) = \Phi(\gamma_{x'}^{-1} * \gamma_{x'x} * \gamma_x), \tag{3.23}$$

where $\gamma_{x'x}$ is the straight line joining x to a neighbouring point x'. Define then the gauge potential $A_\mu(x)$ as:

$$A_\mu(x) = -\frac{i}{g} \lim_{\Delta \to 0} \{h(x', x) - 1\} \tag{3.24}$$

for

$$x'^\nu = x^\nu + \Delta \delta_\mu^\nu. \tag{3.25}$$

Using the composition law as illustrated in Figure 8, one shows that any $\Phi(C)$ can indeed be written in terms of this $A_\mu(x)$ in the usual manner. The recovery of the gauge potential from the loop variables $F_\mu[\xi|s]$ is then complete.

The above construction of $A_\mu(x)$ depends on the choice of the straight line γ_x for each space time point x. This choice is not unique and can be
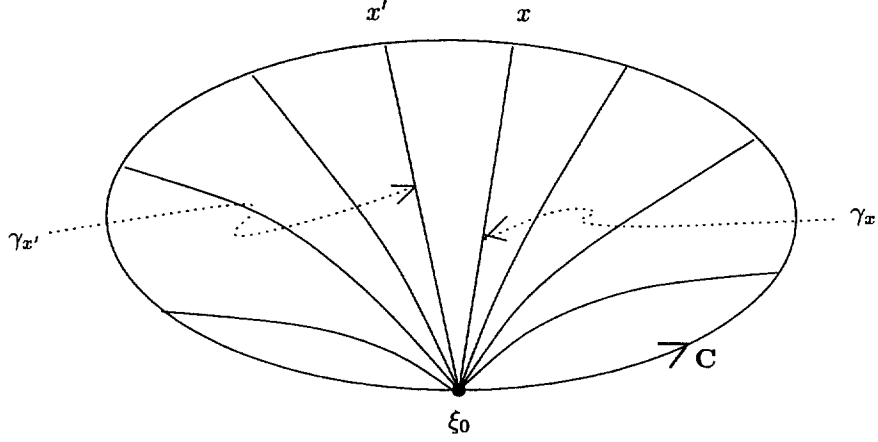
Figure 8: Construction of $\Phi(C)$ from $A_\mu(x)$

replaced by any other choice of an open path for γ_x linking the reference point to x so long as neighbouring points are assigned neighbouring paths. A different choice for the paths γ_x, however, is seen to correspond just to a gauge transformation on $A_\mu(x)$ [9, 7].

4 Electric-Magnetic Duality and its Nonabelian Generalization

In vacuo, the field tensor $F_{\mu\nu}$ in Maxwell theory satisfies the equations:

$$\partial_\nu F^{\mu\nu}(x) = 0, \tag{4.1}$$

and:

$$\partial_\nu {}^*F^{\mu\nu}(x) = 0, \tag{4.2}$$

which are symmetric under the *-operation interchanging electricity with magnetism. This property of the theory is what is loosely known as electric-magnetic duality. In particular, just as the equation (4.2) guarantees that the field $F_{\mu\nu}$ is derivable from a potential $A_\mu(x)$,

$$F_{\mu\nu}(x) = \partial_\nu A_\mu(x) - \partial_\mu A_\nu(x), \tag{4.3}$$

so also the equation (4.1) implies that there is a local 'dual' potential $\tilde{A}_\mu(x)$ such that:

$$ {}^*F_{\mu\nu}(x) = \partial_\nu \tilde{A}_\mu(x) - \partial_\mu \tilde{A}_\nu(x). \tag{4.4}$$

And just as $F_{\mu\nu}$ is invariant under the gauge transformation:

$$A_\mu(x) \longrightarrow A_\mu(x) + \partial_\mu \alpha(x), \tag{4.5}$$

20

for an arbitrary $\alpha(x)$, so is also $^*F_{\mu\nu}$ invariant under the gauge transformation:

$$\tilde{A}_\mu(x) \longrightarrow \tilde{A}_\mu(x) + \partial_\mu \tilde{\alpha}(x), \tag{4.6}$$

for another arbitrary $\tilde{\alpha}(x)$ which need have nothing to do with the $\alpha(x)$ in (4.5). Hence, it follows that the theory is itself invariant under a doubled $U(1) \times \tilde{U}(1)$ gauge transformation where, apart from having the opposite parity (because of the *), the 'magnetic' $\tilde{U}(1)$ as a group is the same as the 'electric' $U(1)$. It is important to note, however, that although there are two independent gauge degrees of freedom as represented by (4.5) and (4.6), the physical degrees of freedom remain that given by either $F_{\mu\nu}$ or $^*F_{\mu\nu}$, but not both, since these two quantities are always related by the algebraic relation (2.8) defining the Hodge star *. The logical steps in the above arguments are summarized in Chart I.

What happens when there are charges around? The Maxwell theory as usually formulated is then not dual symmetric because there are only electric but no magnetic charges. This is, however, merely a whim of nature; the theory itself still keeps the dual symmetry for there is in prinicple nothing (apart from experiment) to stop one introducing magnetic charges into the theory and write for the equations of motion:

$$\partial^\nu F_{\mu\nu}(x) = j_\mu(x), \tag{4.7}$$

and:

$$\partial^{\nu *}F_{\mu\nu}(x) = \tilde{j}_\mu x, \tag{4.8}$$

with j_μ and $\tilde{j}_\mu$ as respectively the electric and magnetic currents. In the standard description in terms of the Maxwell field $F_{\mu\nu}$ and A_μ, electric charges appear as sources of the field as per (4.7), while magnetic charges, as Dirac has taught us and as we have explained in Section 2, will appear as monopoles or topological obstructions in A_μ. However, if one chooses to describe Maxwell theory in terms of the dual fields $^*F_{\mu\nu}$ and $\tilde{A}_\mu$, magnetic charges instead will appear as sources as per (4.8) while electric charges will appear as monopoles. We can thus apply the Wu-Yang criterion to either electric or magnetic charges to derive their equations of motion. In particular, one sees that the standard Lorentz and Dirac equations for electric charges can be deduced as consequences of their topology when regarded as monopoles. The logical structure of duality for Maxwell theory with charges is summarized in Chart III.

The question now is whether the above derivation is generalizable to nonabelian Yang-Mills theory. It is easy to see that a naive extension with the same *-operation as dual transform would fail. Although the field tensor satisfies in vacuo:

$$D^\nu F_{\mu\nu}(x) = 0, \tag{4.9}$$

21

and, if $F_{\mu\nu}$ is derivable from a potential A_μ, also the Bianchi identity:

$$D^{\nu}{}^*F_{\mu\nu}(x) = 0, \qquad (4.10)$$

these two equations, despite appearances, are not dual symmetric. The covariant derivative D_μ occurring in both (4.9) and (4.10) involves the potential A_μ of $F_{\mu\nu}$, whereas for the second equation to be dual to the first, it ought instead to involve a potential, say $\tilde{A}_\mu$, bearing the same relation to the dual field ${}^*F_{\mu\nu}$ as A_μ to $F_{\mu\nu}$, namely:

$$ {}^*F_{\mu\nu}(x) = \partial_\nu \tilde{A}_\mu(x) - \partial_\mu \tilde{A}_\nu(x) + i\tilde{g}\,[\tilde{A}_\mu(x), \tilde{A}_\nu(x)], \qquad (4.11) $$

and this $\tilde{A}(x)$ has no reason to be the same as A_μ. Indeed, one does not even know whether such an $\tilde{A}_\mu$ exists. In contrast to the abelian theory where the equation of motion, namely the Maxwell equation (4.1), implies by the Poincaré Lemma that a potential exists for ${}^*F_{\mu\nu}$, the same is not true for the the nonabelian case. The equation of motion, which is here the Yang-Mills equation (4.9), does not imply a potential for the dual field ${}^*F_{\mu\nu}$. Worse in fact, since Gu and Yang [19] have exhibited many explicit counter-examples of solutions to the equation (4.9) for which no potential for ${}^*F$ exists.

Does that mean then that electric-magnetic duality is not generalizable to Yang-Mills theory? Not necessarily, for it can be that if one defines the dual transform in a different way from *, then duality is retrieved. Supposing this to be true, let us first try to imagine what sort of properties such a generalized dual transform will need to possess. We would want, of course, the new transform to reduce to the Hodge star for the abelian theory as a special case, and to be reversible apart for a sign, like the Hodge star: ${}^*({}^*F) = -F$, so as to qualify as a 'duality'. But the most difficult property to satisfy is that discussed in the preceding paragraph of the existence of a dual potential $\tilde{A}_\mu$ and in order to recover that, we turn back to the abelian theory for inspiration. The reason why abelian duality works with the Hodge star lies in the fact that the source-free Maxwell equation (4.1) implies that there are no monoples in the dual field ${}^*F$, which is in physical terms exactly the condition required by the Poincaré Lemma for the existence of a gauge potential. Thus it appears that a crucial property of the new transform we seek is that the dual field should be so defined as to make the source-free Yang-Mills equation (4.9) equivalent to the statement that there are no monopole charges in the dual field. Given that the occurrence or otherwise of monopole charges in a nonabelian field was shown in our previous sections to be best expressed in loop space language, it is indicated that the same language be adopted also for constructing the generalized dual transform.

These, then, are the leads which set us off to look for a generalized dual transform in loop space [20]. To avoid obscuring the basic arguments with

details, we shall begin with an outline of the construction and only return later to consider the 'proof' of its validity as given in [20].

First, in place of the $F_\mu[\xi|s]$ that we have employed up to now as variables in loop space, we introduce a new set:

$$E_\mu[\xi|s] = \Phi_\xi(s,0)F_\mu[\xi|s]\Phi_\xi^{-1}(s,0), \qquad (4.12)$$

with $\Phi_\xi(s,0)$ defined as in (1.15). These $E_\mu[\xi|s]$ are not gauge invariant like $F_\mu[\xi|s]$ and may not be as useful in general but seem more convenient for dealing with duality. In particular, the condition that there be no monopole charge anywhere in space which, according to Section 3, is expressible in terms of $F_\mu[\xi|s]$ as the vanishing of the loop curvature $G_{\mu\nu}[\xi|s]$, is given in terms of $E_\mu[\xi|s]$ simply as:

$$\delta_\nu(s)E_\mu[\xi|s] - \delta_\mu(s)E_\nu[\xi|s] = 0. \qquad (4.13)$$

On the other hand, the source-free condition (4.9) for Yang-Mills fields which was expressed by Polyakov [8] as the vanishing of the loop divergence of $F_\mu[\xi|s]$ (3.17) remains simply:

$$\delta^\mu(s)E_\mu[\xi|s] = 0 \qquad (4.14)$$

in terms of $E_\mu[\xi|s]$.

Using these new variables $E_\mu[\xi|s]$ for the field, we now define their 'dual' $\tilde{E}_\mu[\eta|t]$ as:

$$\omega^{-1}(\eta(t))\tilde{E}_\mu[\eta|t]\omega(\eta(t)) \qquad (4.15)$$

$$= -\frac{2}{N}\epsilon_{\mu\nu\rho\sigma}\dot{\eta}^\nu(t)\int\delta\xi ds\, E^\rho[\xi|s]\dot{\xi}^\sigma(s)\dot{\xi}^{-2}(s)\delta(\xi(s)-\eta(t)), \quad (4.16)$$

where $\omega(x)$ is a (local) rotation matrix tranforming from the frame in which the orientation in internal symmetry space of the fields $E_\mu[\xi|s]$ are measured to the frame in which the dual fields $\tilde{E}_\nu[\eta|t]$ are measured. The dual transform $\tilde{}$ is by construction reversible apart from a sign, meaning that $\tilde{\tilde{E}} = -E$, and also reduces to the Hodge star for the abelian theory, as required. Further, it can be shown by differentiating (4.16) that:

$$\omega^{-1}(\eta(t))\{\delta_\nu(t)\tilde{E}_\mu[\eta|t] - \delta_\mu(t)\tilde{E}_\nu[\eta|t]\}\omega(\eta(t)) =$$
$$-\frac{1}{N}\int\delta\xi ds\epsilon_{\mu\nu\rho\sigma}\{\dot{\eta}^\beta(t)\dot{\xi}^\alpha(s)-\dot{\eta}^\alpha(t)\dot{\xi}^\beta(s)\}\delta_\rho(s)E^\rho[\xi|s]\dot{\xi}^{-2}(s)\delta(\xi(s)-\eta(t)). (4.17)$$

Now equation (4.17) means that so long as the Yang-Mills field is source-free and therefore (4.14) is satisfied, then

$$\delta_\nu(t)\tilde{E}_\mu[\eta|t] - \delta_\mu(t)\tilde{E}_\nu[\eta|t] = 0, \qquad (4.18)$$

23

or that there are no monopoles in the dual field. Hence, by the Extended Poincaré Lemma discussed in Section 3, we deduce that a dual potential $\tilde{A}_\mu(x)$ will exist in that case.

With this generalized dual transform in hand, we can now analyse the dual structure of nonabelian Yang-Mills theory. We shall begin with the pure theory with no source either in E or $\tilde{E}$, for which the structure is summarized in Chart III. Consider first the left half. The field action is $\mathcal{A}^0$ which can be expressed either in terms of the gauge potential $A_\mu(x)$ via the field tensor $F_{\mu\nu}(x)$, or in terms of the loop variable $E_\mu[\xi|s]$. Adopting $A_\mu(x)$ as variables, the theory can be developed along conventional lines. Extremizing $\mathcal{A}^0$ with respect to $A_\mu(x)$, one obtains the standard Yang-Mills equation (4.9), which can be written also in terms of $E_\mu[\xi|s]$ as (4.14). By (4.17), however, this is equivalent to (4.18) and implies the existence of a dual gauge potential $\tilde{A}_\mu(x)$. This chain of arguments is summarized in the left-most column of Chart III. Alternatively, adopting $E_\mu[\xi|s]$ as variables which are redundant, the constraint (4.13) has to be imposed, which is incorporated via Lagrange multipliers $W_{\mu\nu}[\xi|s]$ into the action $\mathcal{A}$. Extremizing now $\mathcal{A}$ with respect to $E_\mu[\xi|s]$, one obtains an equation relating $E_\mu[\xi|s]$ to the Lagrange multipliers, from which relation the dual gauge potential $\tilde{A}_\mu(x)$ can then be constructed. This chain of arguments is summarized in the second column of Chart III. Furthermore, the existence of the dual potential under gauge transformation of which the theory is invariant, together with the original invariance under gauge transformations of $A_\mu(x)$, implies that the theory has overall a doubled gauge symmetry $SU(N) \times \widetilde{SU}(N)$.

The dual tranform (4.16) allows one to express the field action $\mathcal{A}^0$ also in terms of the dual loop variables $\tilde{E}_\mu[\xi|s]$. The fact now that the dual transform is reversible means that in going over into the dual representation, the steps outlined in the preceding paragraph can all be repeated with only occasional changes in signs as shown on the right half of Chart III. Comparing this Chart with Chart I shows a very close analogy with the abelian theory.

Next, we turn to Yang-Mills theory with charges. In this case, one works exclusively with loop variables which admit an easier implementation of the Wu-Yang criterion. Consider first again the left-half of Chart IV. The free field-particle system as represented by the free action $\mathcal{A}^0$ is subjected to the constraint:

$$\delta_\nu(s)E_\mu[\xi|s] - \delta_\mu(s)E_\nu[\xi|s] = -4\pi J_{\mu\nu}[\xi|s], \qquad (4.19)$$

which is both the constraint imposed by the Wu-Yang criterion for deriving the dynamics of the monopoles and that required by the Extended Poincaré Lemma of Section 3 for removing the redundancy in the loop variables. Whether one is dealing with a classical or a Dirac particle is distinguished by the choice of the free particle action and the form of the current $J_{\mu\nu}[\xi|s]$,

as shown in the separate columns of Chart IV. Incorporating the constraint (4.19) into the action $\mathcal{A}$, and extremizing with respect to the field variables $E_\mu[\xi|s]$ and particle variables $Y^\mu(\tau)$ or $\psi(x)$, one obtains the equations of motion. One notes in particular the equation for the Dirac monopole, which couples the monopole to the dual potential $\tilde{A}(x)$ and takes a form exactly dual of that for a 'colour' charge. That the monopole is so coupled to the field confirms that the dual potential constructed does play the expected role of a connection giving parallel phase transport for the monopole wave function. It also implies that the theory has an $\widetilde{SU}(N)$ symmetry corresponding to the phase of monopole wave functions in addition to the original $SU(N)$ gauge symmetry corresponding to the phase of the wave functions of 'colour' charges, giving the theory thus in all an $SU(N) \times \widetilde{SU}(N)$ gauge symmetry. The equation for the classical monopole is perhaps less familiar but is also exactly dual to the so-called Wong equation [21] for a classical 'colour' charge. Again, given that the dual transform is reversible, one obtains a near symmetry apart from signs between the left and right halves of the Chart.

This then completes our outline of nonabelian duality barring the derivations of some formulae that we have used, which, unfortunately, will involve some rather delicate operations in loop space, and are as yet far from rigorous. Even more than before, the lack of a general loop calculus, already mentioned in Section 1, is here strongly felt. We shall not go through all the details which the reader can find in our original publications [20, 22], but shall just pick up a few representative points for illustration.

First, let us return to the variables $E_\mu[\xi|s]$ defined in (4.12). Recalling the definition (1.8) of the Polyakov variable $F_\mu[\xi|s]$ one sees that $E_\mu[\xi|s]$ can be pictured as the bold curve in Figure 9 where the phase factors $\Phi_\xi(s,0)$ in (4.12) have cancelled parts of the faint curve representing $F_\mu[\xi|s]$. In contrast to $F_\mu[\xi|s]$, therefore, $E_\mu[\xi|s]$ depends really only on a "segment" of the loop ξ from s_- to s_+. Notice that the δ-function $\delta(s - s')$ inherent in our definition (1.6) and (1.7) of the loop derivative $\delta_\mu(s)$ is represented in the figure as a bump function centred at s with width $\epsilon = s_+ - s_-$. The reason for doing so is that, as usual in most functional formulations, our treatment here involves operations with the δ-function which need to be "regularized" to be given a meaning. Our procedure is to take first the δ-function as a bump function with finite width ϵ and height h, and afterwards take the zero width limit. In (1.6) and (1.7), the limit $\epsilon \to 0$ with $\Delta = \epsilon h$ held fixed is taken first, to be followed by the limit $h \to 0$.

For example, suppose we wish to take the loop derivative $\delta_\nu(s)$ of the quantity $E_\mu[\xi|s]$ at the same value of s. Clearly, a loop derivative has a meaning only if there is a segment of the loop on which it can operate. Therefore, to define this derivative, we shall first regard $E_\mu[\xi|s]$ as a segmental

25
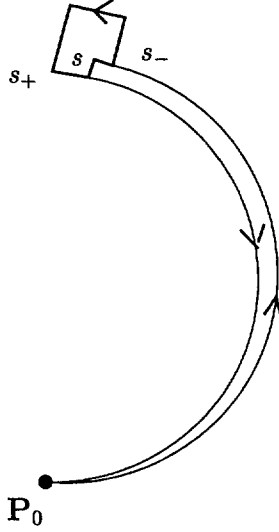
Figure 9: Illustration for $E_\mu[\xi|s]$

quantity dependent on the segment of the loop ξ from $s - \epsilon/2$ to $s + \epsilon/2$. We then define the loop derivative $\delta_\nu(s)$ using the normal procedure on this segment, and afterwards take the limit $\epsilon \to 0$. Following this procedure, we have then by (4.12) and (1.8):

$$\delta_\nu(s')E_\mu[\xi|s] = \Phi_\xi(s,0)\{\delta_\nu(s')F_\mu[\xi|s] + ig\theta(s-s')[F_\nu[\xi|s'], F_\mu[\xi|s]]\}\Phi_\xi^{-1}(s,0), \tag{4.20}$$

where $\theta(s)$ is the Heaviside θ-function, so that:

$$G_{\mu\nu}[\xi|s] = \Phi_\xi^{-1}(s,0)\{\delta_\nu(s)E_\mu[\xi|s] - \delta_\mu(s)E_\nu[\xi|s]\}\Phi_\xi(s,0). \tag{4.21}$$

This shows that the absence of monopoles, which was given in terms of the variables $F_\mu[\xi|s]$ by the vanishing of the loop curvature $G_{\mu\nu}[\xi|s]$ is here translated in terms of $E_\mu[\xi|s]$ to read as the vanishing of the 'curl' of $E_\mu[\xi|s]$ as was claimed in (4.13) above.

Next, as another example, let us check that our generalized dual transform (4.16) does reduce to the Hodge star when the theory is abelian. To see this, we let the segmental width of $\tilde{E}_\mu[\eta|t]$ in (4.16) go to zero so that we can use the formula given in (1.14) for $F_\mu[\xi|s]$ to write the left-hand side in terms of local quantities:

$$\omega^{-1}(x)\tilde{F}_{\mu\nu}(x)\omega(x) = -\frac{2}{\bar{N}}\epsilon_{\mu\nu\rho\sigma}\int\delta\xi ds\, E^\rho[\xi|s]\dot{\xi}^\sigma(s)\dot{\xi}^{-2}(s)\delta(x - \xi(s)). \tag{4.22}$$

To evaluate the right-hand side, we recall that our procedure is to do the integral before taking the width of the segment in $E_\mu[\xi|s]$ to zero. In other words, within the integral, the loop ξ can still vary by a δ-functional bump

26
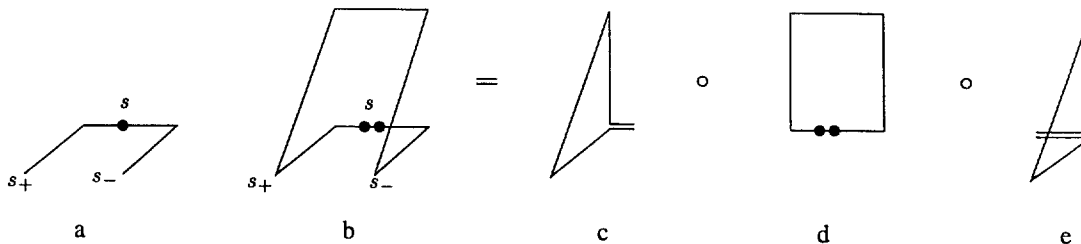
Figure 10: Illustration for the Integrand in Dual Transform

as illustrated in Figure 10 (a). For such a ξ, $E_\mu[\xi|s]$, which is obtained by making a δ-functional variation along the direction μ, will take on the shape depicted in Figure 10 (b). This last figure can be expressed as the product of three factors, namely Figures 10(c),(d),(e) in the order indicated. In the abelian theory, the ordering of the factors is unimportant so that the factors of Figures (c) and (e) cancel in the limit when the segmental width $\epsilon \to 0$, leaving only the factor of Figure (d), which can as before be expressed as $F_{\mu\alpha}(\xi(s))\dot{\xi}^\alpha(s)$, giving:

$$
\begin{aligned}
\tilde{F}_{\mu\nu}(x) &= -\frac{2}{\bar{N}}\epsilon_{\mu\nu\rho\sigma} \int \delta\xi ds\, F^{\rho\alpha}(\xi(s))\dot{\xi}_\alpha(s)\dot{\xi}^\sigma(s)\dot{\xi}^{-2}(s)\delta(x - \xi(s)) \\
&= -\tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}(x).
\end{aligned}
\tag{4.23}
$$

This is just the Hodge star relation if we identify $\tilde{F}_{\mu\nu}(x)$ with ${}^*F_{\mu\nu}(x)$. On the other hand, for a nonabelian theory, the factors of Figures 10 (c) and (e) cannot be commuted through the factor of Figure (d) so that the above reduction to the Hodge star relation will not go through. This last statement is important for otherwise our claim of nonabelian duality under the generalized dual transform (4.16) would be in contradiction with the result of Gu and Yang quoted above [19].

These two examples, we hope, should give a taste of the sort of arguments one had to go through to justify the results claimed above. They serve to illustrate the level of rigour one is able at present to achieve, which is unfortunately not as high as one could wish. Except for this reservation, one could claim that a nonabelian generalization of electric-magnetic duality is now established.

5 't Hooft's Order-Disorder Parameters

As in electromagnetism, the dual symmetry in nonabelian Yang-Mills theory is established only for classical fields. In contrast to electromagnetism,

27

however, where the classical theory has already a wide range of applications, Yang-Mills theory applied to physics involves almost always the quantum theory. The construction of a quantized version of the above results looks difficult and only the most tentative of beginnings of an attempt have so far been made [23]. However, one very useful result for the quantum theory has already been derived which is the subject of this section.

In his famous study of the confinement problem in nonabelian gauge theories, 't Hooft [24] introduced 2 loop-dependent operators $A(C)$ and $B(C)$ which satisfy the following commutation relation:

$$A(C)B(C') = B(C')A(C)\exp(2\pi i l/N) \qquad (5.1)$$

for $su(N)$ symmetry and any 2 spatial loops C and C' with linking number l between them. $A(C)$ is given explicitly as:

$$A(C) = \text{Tr}\left[P\exp ig\oint_C A_i(x)dx^i\right], \qquad (5.2)$$

and in the words of 't Hooft measures the magnetic flux through C while creating electric flux along C. On the other hand, $B(C)$ measures the electric flux through C while creating magnetic flux along C, and plays thus an exactly dual role to $A(C)$. For lack of a dual potential, however, $B(C)$ was not given a similar explicit expression to (5.2).

Now, in the preceding section, one claims that there does indeed exist a dual potential $\tilde{A}_\mu$ for nonabelian gauge fields, although by duality is now meant no longer the Hodge star but a more complicated transform which reduces to the Hodge star only for the abelian theory. That being the case, one ought to have:

$$B(C) = \text{Tr}\left[P\exp i\tilde{g}\oint_C \tilde{A}_i(x)dx^i\right] \qquad (5.3)$$

as the explicit expression for $B(C)$, with the dual coupling $\tilde{g}$ related to g by a Dirac quantization condition. In other words, starting from the formulae (5.2) and (5.3) with A_μ nd $\tilde{A}_\mu$ related in the manner detailed in the preceding section, one ought to be able to deduce the commutation relation (5.1) required by 't Hooft. The success in doing so would be an important check on the consistency of the proposed framework for duality. It also means that the duality as defined above accords with what 't Hooft called duality so that the important results he derived are applicable to the present case.

That (5.1) is indeed satisfied for (5.2) and (5.3) is shown in [25]. To appreciate easier how this obtains, recall first how the parallel assertion to (5.1) can be deduced in the abelian case:

$$\left[ie\oint_C A_i(x)dx^i,\ i\tilde{e}\oint_{C'}\tilde{A}_i(x)dx^i\right] = 2\pi l. \qquad (5.4)$$

28

Using Stokes' theorem, the second integral over C' in (5.4) can be written as a surface integral, thus:

$$-i\tilde{e} \int\int_{\Sigma_{C'}} {}^*F_{ij}\, d\sigma^{ij}, \qquad (5.5)$$

or, by the definition of the Hodge star, in terms of the electric field strength $\mathcal{E}_i = F_{0i}$ as:

$$i\tilde{e} \int\int_{\Sigma_{C'}} \mathcal{E}_i\, d\sigma^i, \qquad (5.6)$$

where $\Sigma_{C'}$ is some surface both spanning and bounded by C'.

Consider first the simple case for linking number 1 between C and C'. The loop C in that case will intersect the surface $\Sigma_{C'}$ at some point x_0. (C may of course intersect $\Sigma_{C'}$ more than once, but the extra intersections occurring pairwise with opposite orientations, their contributions to the commutator will all cancel, leaving in effect just one intersection.) Except at this point x_0, all points on C are spatially separated from points on $\Sigma_{C'}$ so that, using the canonical commutation relation between $A_i(x)$ and $\mathcal{E}_i(x)$:

$$[\mathcal{E}_i(x), A_j(x')] = i\,\delta_{ij}\,\delta(x - x') \qquad (5.7)$$

we have:

$$\left[ie \oint_C A_i(x)dx^i, i\tilde{e} \int\int_{\Sigma_{C'}} \mathcal{E}_j d\sigma^j \right] = ie\tilde{e}, \qquad (5.8)$$

which by the Dirac quantization condition (2.5)(see footnote for the rationalized couplings adopted here) gives the answer (5.4) for $l = 1$ as required. In case C' winds around C more than once, say l times, then C will intersect $\Sigma_{C'}$ at effectively l points for each of which the above applies, so that (5.4) still remains valid.

What happens when we generalize to the nonabelian case? Then $A(C)$ and $B(C)$ are each a trace of an ordered product of noncommuting factors, *not*, in spite of appearances (due to the somewhat misleading standard notation), an exponential of a line integral for which Stokes' Theorem applies, so that the above arguments no longer work. Nevertheless, one finds that one may still associate with each a surface in an analogous fashion [25]. Take $B(C')$, for example. The phase factor:

$$\tilde{\Phi}(C') = P \exp i\tilde{g} \oint_{C'} \tilde{A}_i dx^i \qquad (5.9)$$

of which $B(C')$ is the trace, can be written, according to [20], as:

$$\tilde{\Phi}(C') = \prod_{t=0\to 2\pi} (1 - i\tilde{g}\tilde{W}[\eta|t]) \sim \prod_{t=0\to 2\pi} \exp(-i\tilde{g}\tilde{W}[\eta|t]), \qquad (5.10)$$

29

where η, for $t = 0 \to 2\pi$, is a parametrization of C', and $\tilde{W}$ is a segmental quantity the 'loop gradient' of which, $\delta_\mu \tilde{W}$, as indicated in Chart III, is the field variable $\tilde{E}_\mu$. One can thus write symbolically:

$$\tilde{W}[\eta|t] = \int_{\eta_0}^{\eta(t)} \delta \eta'^\nu(t)\, \tilde{E}_\nu[\eta'|t], \tag{5.11}$$

where the integral in (5.11) denotes a 'segmental' integral along some path from a reference point η_0 to the point $\eta(t)$, so that in ordinary space, this path appears as a ribbon. Piecing such ribbons together as $\eta(t)$ moves along C' in (5.10), one obtains a surface $\Sigma_{C'}$ spanning over and bounded by C' as suggested. In as much as the reference point η_0 and the path joining it to $\eta(t)$ are both arbitrary for (5.11) to hold, one can choose $\Sigma_{C'}$ to be completely space-like. This surface will again intersect the loop C of $A(C)$ at some point x_0. (Previous remarks in the abelian case about multiple intersections and higher linking numbers between C and C' will still apply and need not be repeated.)

The remaining arguments for deriving (5.1) [25] then follow along much the same lines as for (5.4) in the abelian case although they are naturally more complicated because of the noncommutative quantities involved. Thus, starting from the formula (5.11), we write $\tilde{E}[\eta|t]$ via the dual transform (4.16) given in the last section in terms of its dual $E[\xi|s]$, which is then related to the usual Yang-Mills local field tensor $F_{\mu\nu}(x)$. The commutation with the operator $A(C)$ using the canonical commutation relation between $A_i^\alpha(x)$ and the 'electric' field strengths $\mathcal{E}_i^\beta(x) = F_{0i}^\beta$ then picks out the one point of intersection x_0 between the loop C and the surface associated to $B(C')$ as described in the last paragraph. Again, the Dirac quantization condition comes in in relating the dual charge $\tilde{g}$ to g giving rise to the factor 2π in the exponent on the right-hand side of (5.1). For details, the reader is referred to the original reference [25].

Although the derivation of (5.1) represents but a small step on the road to quantizing the classical formalism developed in the previous sections, it is of much practical importance in allowing one to apply 't Hooft's far-reaching results [24], which have been used to good effect in the Dualized Standard Model [1]. Furthermore, as a by-product, one has obtained in (5.3) an explicit formula for the operator $B(C)$ which may be useful in the problem of confinement. Although other, and presumably equivalent, explicit formulae for $B(C)$ have previously been suggested [26], they are, in contrast to (5.3), usually given as dependent not only on the loop C but also on the particular surface spanning it.

6 Concluding Remarks

Starting from a loop space formulation of Yang-Mills theory which helps to clarify the role that monopoles play in defining the dynamics and in guaranteeing the existence of gauge potentials, a generalization of electric-magnetic duality to the nonabelian theory is derived. The result implies in particular that the gauge symmetry is doubled, say from $SU(N)$ to $SU(N) \times \widetilde{SU}(N)$, where the second factor has opposite parity to the first. However, the physical degrees of freedom, of course, remain the same, and the theory can be described in terms of either the usual Yang-Mills potential $A_\mu(x)$ or a dual potential $\tilde{A}_\mu(x)$. The theory then admits both nonabelian 'electric' and nonabelian 'magnetic' charges, where the former appear as sources of A_μ but as monopoles of $\tilde{A}_\mu$, while the latter appear as monopoles of A_μ but sources of $\tilde{A}_\mu$.

Although these results have been derived only for classical fields, one result is known for the quantum theory, namely that the Dirac phase factors (or Wilson loops) constructed out of A_μ and $\tilde{A}_\mu$ satisfy the 't Hooft commutation relations, so that his results apply. Hence one concludes, in particular, that if $SU(N)$ is in the confined phase then its dual $\widetilde{SU}(N)$ is in the Higgs phase, and vice versa.

When applied to the Standard Model with symmetry $su(3) \times su(2) \times u(1)$, one concludes first that each symmetry has its dual and particles can in principle carry an 'electric' as well as a 'magnetic' charge of each symmetry. Secondly, using the corollary stated above to 't Hooft's result, one concludes that since colour $su(3)$ is confined, dual colour $\widetilde{su}(3)$ is broken, and since electroweak $su(2)$ is in the Higgs phase, dual electroweak $\widetilde{su}(2)$ is in the confined phase. If we assume that these effects have correspondence in Nature, then their manifestations should lead to very interesting consequences. A possible scenario for the physical realization of these effects is the subject of our companion paper [1].

Acknowledgement

31

Chart I: Pure Electromagnetism.

$U(1)$ invariance $(A_\mu \to A_\mu + \partial_\mu \Lambda)$

$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$ (F derivable from potential A)

$\mathcal{A}^0 = -\frac{1}{16\pi} \int (\partial_\nu A_\mu - \partial_\mu A_\nu)^2$ extremize w.r.t. A_μ

$\partial_\nu F^{\mu\nu} = 0$ (no el. charge)

$\mathcal{A}^0 = -\frac{1}{16\pi} \int F_{\mu\nu} F^{\mu\nu}$

$\mathcal{A} = -\frac{1}{16\pi} \int F_{\mu\nu} F^{\mu\nu} + \int \lambda_\mu (\partial_\nu {}^*F^{\mu\nu})$ extremize w.r.t. $F_{\mu\nu}$

$\partial_\nu {}^*F^{\mu\nu} = 0$ (no mag. charge)

$F_{\mu\nu} = 2\pi\epsilon_{\mu\nu\rho\sigma}(\partial^\rho \lambda^\sigma - \partial^\sigma \lambda^\rho)$

${}^*F_{\mu\nu} = \partial_\nu \bar{A}_\mu - \partial_\mu \bar{A}_\nu$
$\bar{A}_\mu = 4\pi\lambda_\mu$
(${}^*F$ derivable from potential $\bar{A}$)

$\widetilde{U(1)}$ invariance $(\bar{A}_\mu \to \bar{A}_\mu + \partial_\mu \bar{\Lambda})$

${}^*F_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}$

$\mathcal{A}^0 = \bar{\mathcal{A}}^0$

$\partial_\nu F^{\mu\nu} = 0$ (no el. charge)

$\bar{\mathcal{A}} = \frac{1}{16\pi} \int {}^*F_{\mu\nu} {}^*F^{\mu\nu} - \int \bar{\lambda}_\mu (\partial_\nu F^{\mu\nu})$ extremize w.r.t. ${}^*F_{\mu\nu}$

${}^*F_{\mu\nu} = -2\pi\epsilon_{\mu\nu\rho\sigma}(\partial^\rho \bar{\lambda}^\sigma - \partial^\sigma \bar{\lambda}^\rho)$

$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$
$A_\mu = 4\pi\bar{\lambda}_\mu$
(F derivable from potential A)

$U(1)$ invariance $(A_\mu \to A_\mu + \partial_\mu \Lambda)$

${}^*F_{\mu\nu} = \partial_\nu \bar{A}_\mu - \partial_\mu \bar{A}_\nu$ (${}^*F$ derivable from potential $\bar{A}$)

$\widetilde{U(1)}$ invariance $(\bar{A}_\mu \to \bar{A}_\mu + \partial_\mu \bar{\Lambda})$

$\bar{\mathcal{A}}^0 = \frac{1}{16\pi} \int (\partial_\nu \bar{A}_\mu - \partial_\mu \bar{A}_\nu)^2$ extremize w.r.t. $\bar{A}_\mu$

$\bar{\mathcal{A}}^0 = \frac{1}{16\pi} \int {}^*F_{\mu\nu} {}^*F^{\mu\nu}$

$\partial_\nu {}^*F^{\mu\nu} = 0$ (no mag. charge)

POINCARE LEMMA

Chart I: Pure Electromagnetism.

MAGNETIC CHARGE

Defining constraint $\partial_\nu {}^*F^{\mu\nu} = -4\pi\tilde{j}^\mu$ (Gauss' law)

POINCARE LEMMA

Potential A for F patched

$\tilde{j}^\mu = \tilde{e}\int d\tau\,\frac{dY^\mu}{d\tau}\delta(x - Y(\tau))$

$A^0 = A_F^0 - m\int d\tau$

$A^0 = A_F^0$

$A_F^0 = -\frac{1}{16\pi}\int F_{\mu\nu}F^{\mu\nu}$

$A = A^0 + \int\lambda_\mu(\partial_\nu {}^*F^{\mu\nu} + 4\pi\tilde{j}^\mu)$ extremize w.r.t. $F_{\mu\nu}, Y^\mu$

$m\frac{d^2Y^\mu}{d\tau^2} = -\tilde{e}\,{}^*F^{\mu\nu}\frac{dY_\nu}{d\tau}$ (dual Lorentz eq.)

δY

$F^{\mu\nu} = 4\pi\epsilon^{\mu\rho\sigma}\partial_\sigma\lambda_\rho$

δF

$\partial^\nu F_{\mu\nu} = 0$ (Maxwell eq.)

POINCARE LEMMA

${}^*F$ derivable from potential $\tilde{A}$

$A = A^0 + \int\lambda_\mu(\partial_\nu {}^*F^{\mu\nu} + 4\pi\tilde{j}^\mu)$ extremize w.r.t. $F_{\mu\nu}, \psi$

$A^0 = A_F^0 + \int\bar\psi(i\slashed\partial - m)\psi$

$\tilde{j}^\mu = \tilde{e}\bar\psi\gamma^\mu\psi$

$\delta\psi$

$(i\slashed\partial - m)\psi = -\tilde{e}\tilde{\slashed{A}}\psi$ $\tilde{A}_\mu = 4\pi\lambda_\mu$ (dual Dirac eq.)

$\widetilde{U(1)}$ invariance $\tilde{A}_\mu \to \tilde{A}_\mu + \partial_\mu\tilde\Lambda$ $\psi \to \exp i\tilde{e}\tilde\Lambda\,\psi$

${}^*F_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$

$\tilde{A}^0 = \tilde{A}_F^0$

ELECTRIC CHARGE

Defining constraint $\partial_\nu F^{\mu\nu} = -4\pi j^\mu$ (Maxwell eq.)

POINCARE LEMMA

Potential $\tilde{A}$ for ${}^*F$ patched

$j^\mu = e\bar\psi\gamma^\mu\psi$

$\tilde{A}^0 = \tilde{A}_F^0 - \int\bar\psi(i\slashed\partial - m)\psi$

$\tilde{A}_F^0 = \frac{1}{16\pi}\int {}^*F_{\mu\nu}{}^*F^{\mu\nu}$

$\tilde{A} = \tilde{A}^0 - \int\tilde\lambda_\mu(\partial_\nu F^{\mu\nu} + 4\pi j^\mu)$ extremize w.r.t. ${}^*F_{\mu\nu}, \psi$

δv

$(i\slashed\partial - m)\psi = -e\slashed{A}\psi$ $A_\mu = 4\pi\lambda_\mu$ (Dirac eq.)

$U(1)$ invariance $A_\mu \to A_\mu + \partial_\mu\Lambda$ $\psi \to \exp ie\Lambda\,\psi$

${}^*F^{\mu\nu} = -4\pi\epsilon^{\mu\nu\rho\sigma}\partial_\sigma\tilde\lambda_\rho$

$\delta {}^*F$

$\partial^\nu {}^*F_{\mu\nu} = 0$ (Gauss' law)

POINCARE LEMMA

F derivable from potential A

$\tilde{A}^0 = \tilde{A}_F^0 + m\int d\tau$

$\tilde{A} = \tilde{A}^0 - \int\tilde\lambda_\mu(\partial_\nu F^{\mu\nu} + 4\pi j^\mu)$ extremize w.r.t. ${}^*F_{\mu\nu}, Y^\mu$

$j^\mu = e\int d\tau\,\frac{dY^\mu}{d\tau}\delta(x - Y(\tau))$

δY

$m\frac{d^2Y^\mu}{d\tau^2} = -e F^{\mu\nu}\frac{dY_\nu}{d\tau}$ (Lorentz eq.)
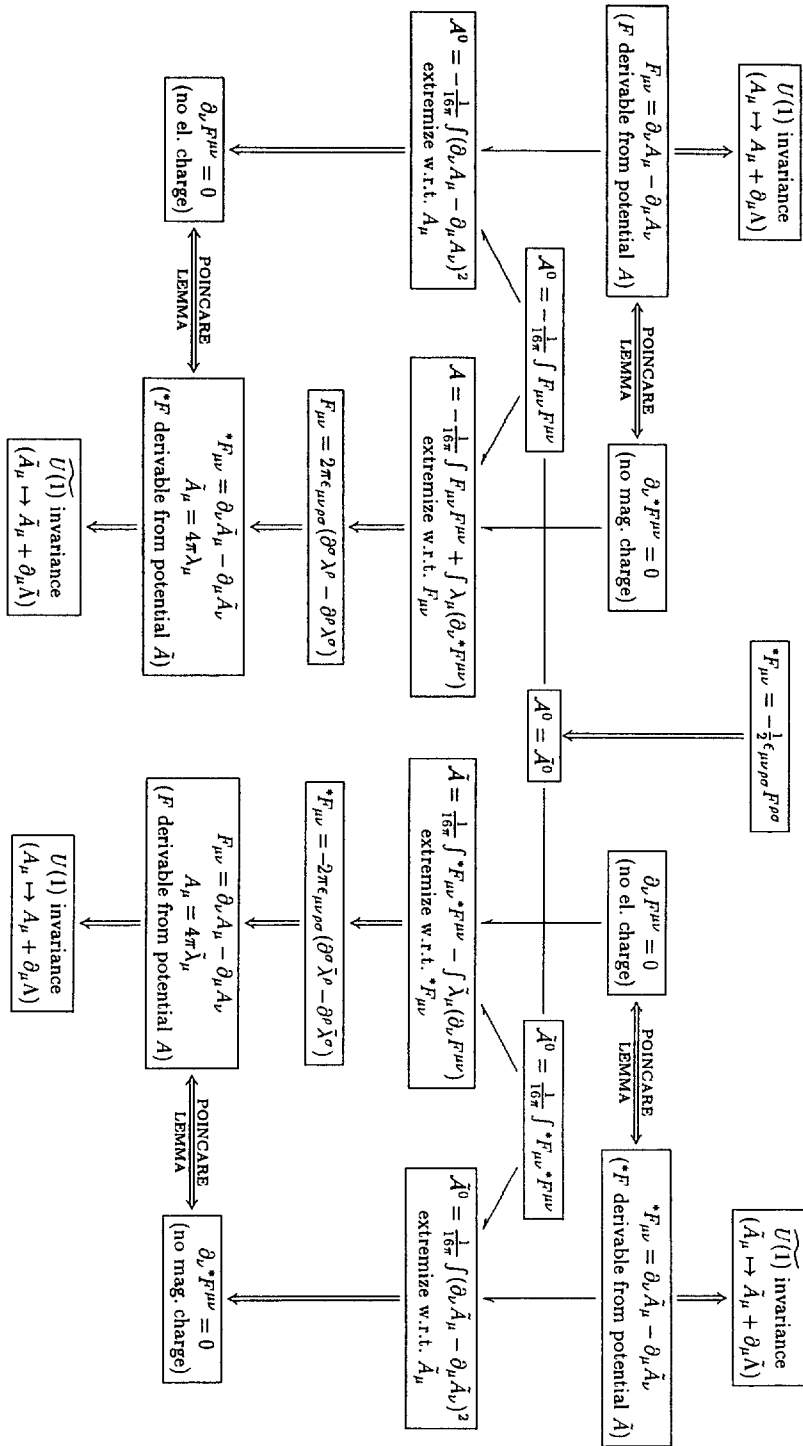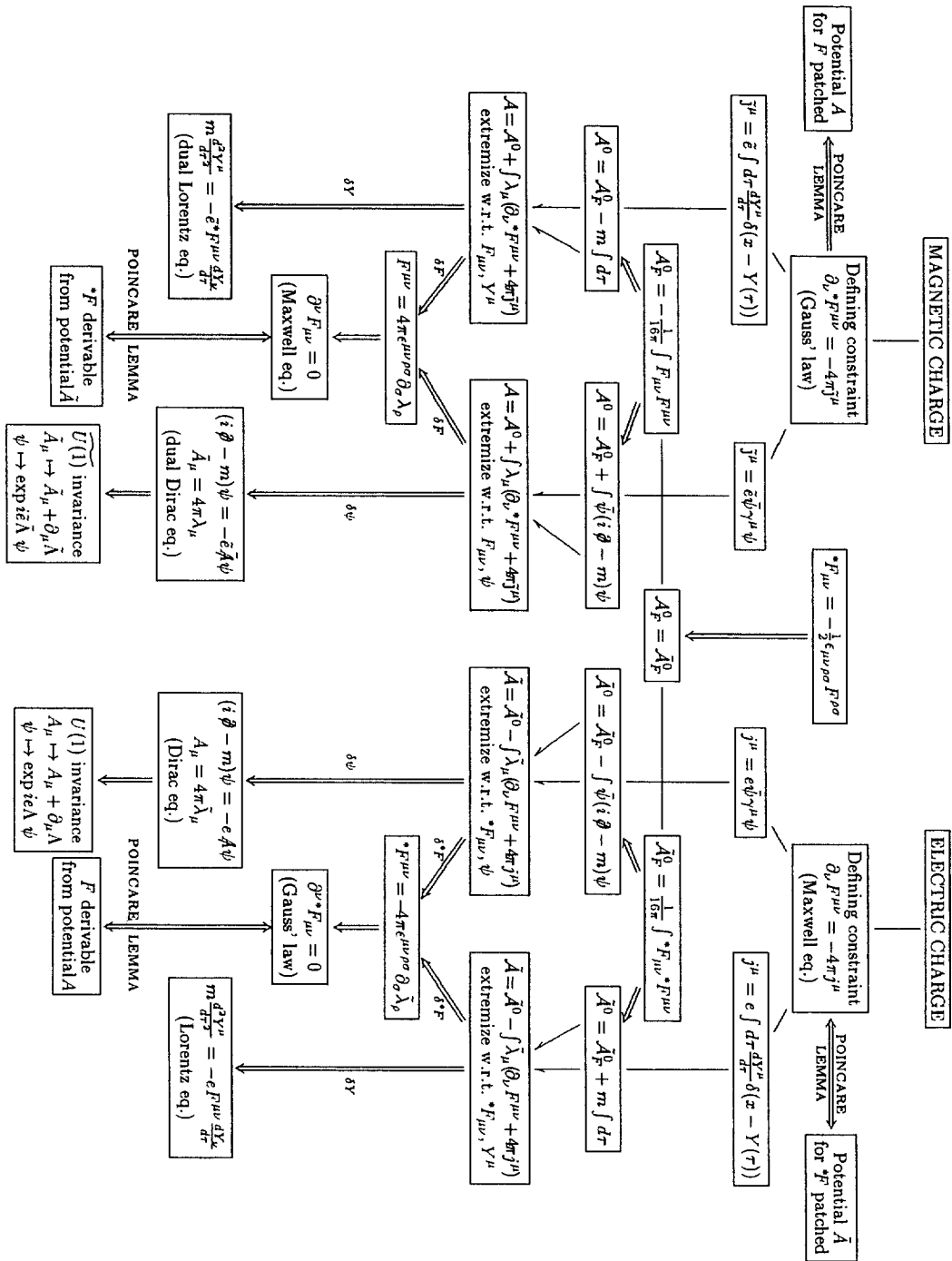
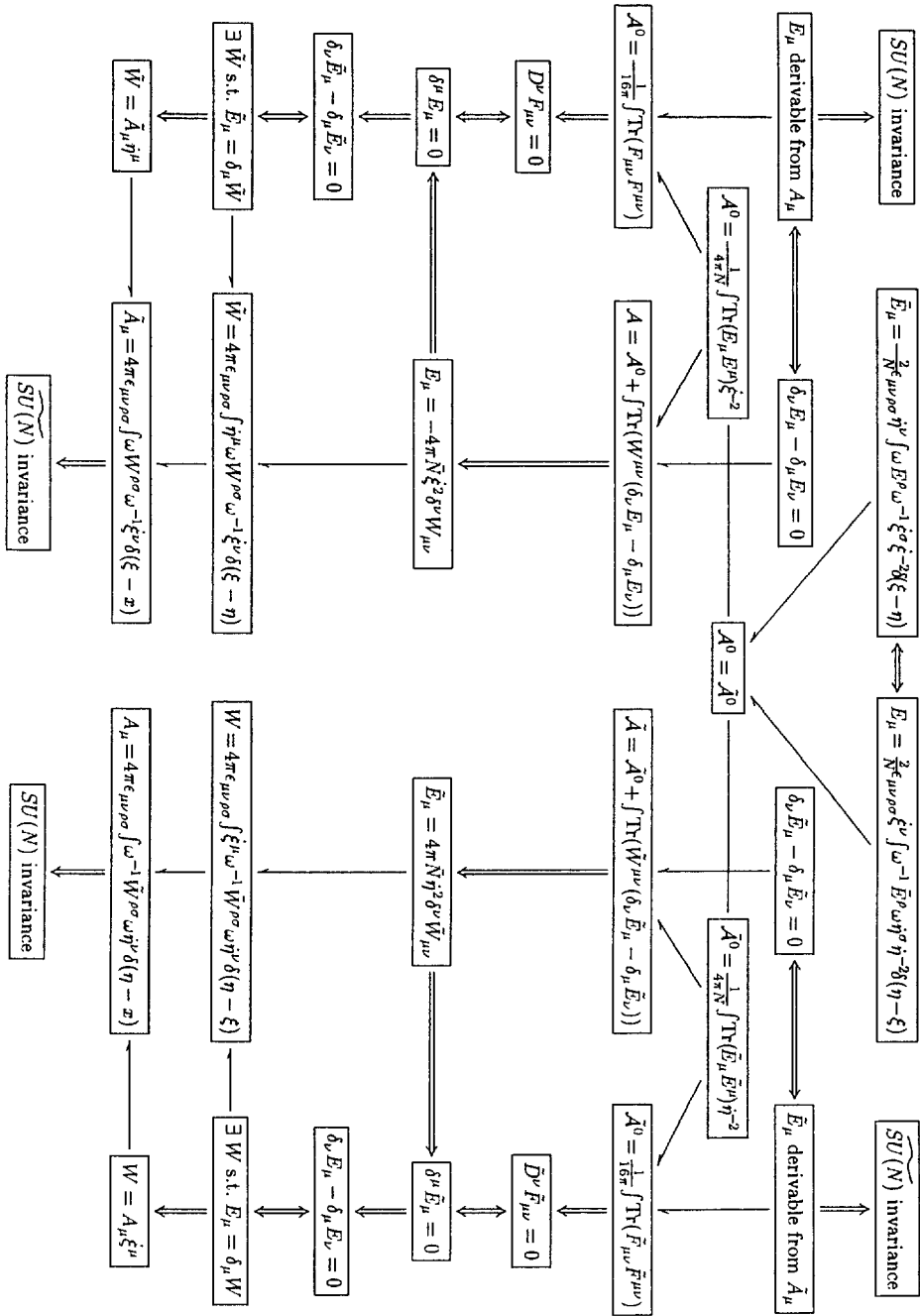Chart II: Electromagnetism with charges.
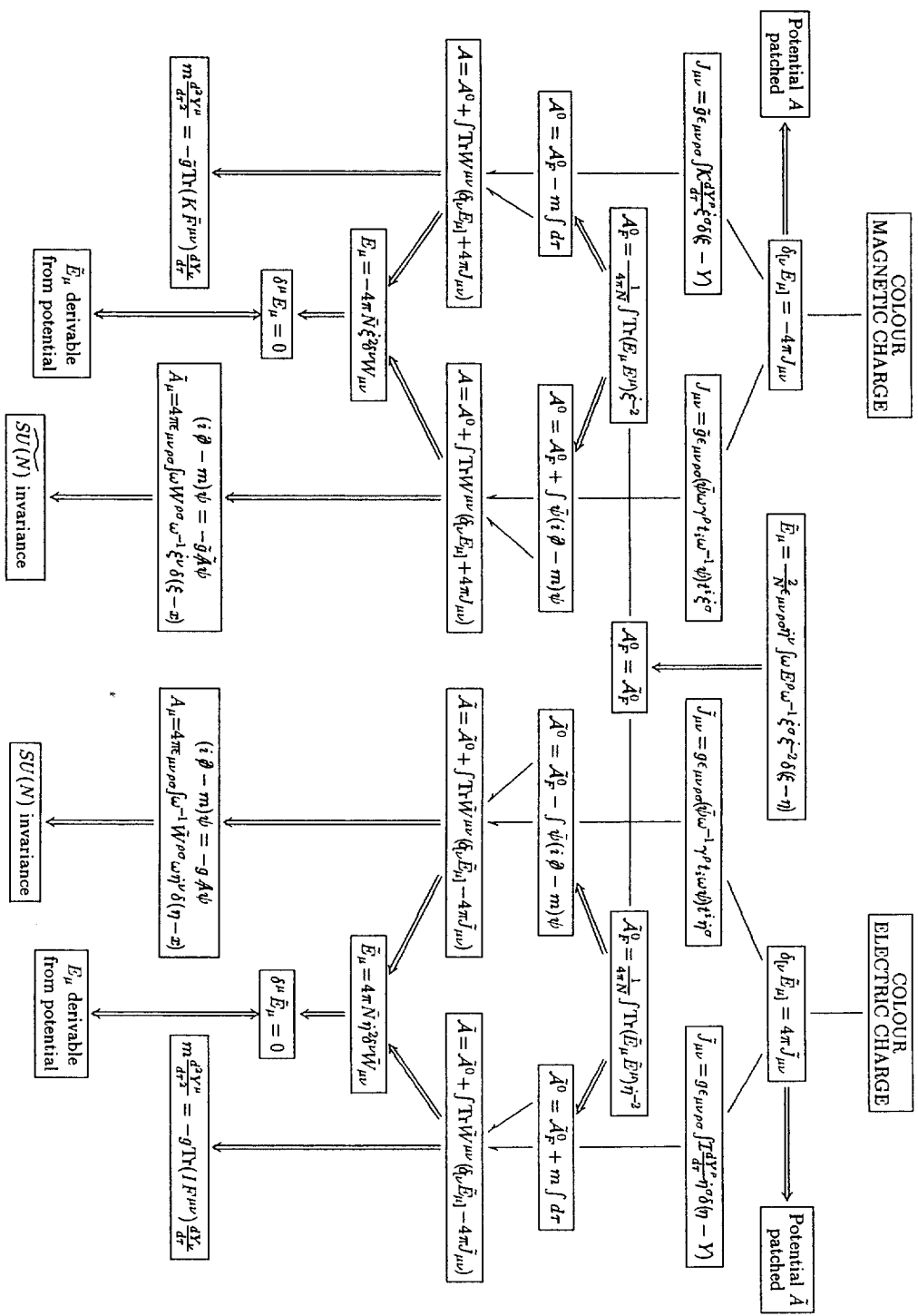
Chart III: Pure Yang–Mills Theory.

34

Chart IV: Yang–Mills with charges.

References

[1] Chan Hong-Mo and Tsou Sheung Tsun, companion paper. For a brief summary see Chan Hong-Mo, José Bordes, and Tsou Sheung Tsun, hep-ph/9809272, talk given at ICHEP'98 in Vancouver, to appear in the conference proceedings.

[2] Y. Aharonov and D. Bohm, Phys. Rev. 115, 485 (1959).

[3] Tai Tsun Wu and Chen Ning Yang, Phys. Rev. D12, 3845 (1975).

[4] Tai Tsun Wu and Chen Ning Yang, Phys. Rev. D12, 3843 (1975).

[5] E. Corrigan and B. Hasslacher, Phys. Lett. 81B, 181, (1979).

[6] Chan Hong-Mo and Tsou Sheung Tsun, Acta Phys. Pol. B17, 259, (1986).

[7] Chan Hong-Mo and Tsou Sheung Tsun, *Some Elementary Gauge Theory Concepts* (World Scientific, Singapore, 1993).

[8] A.M. Polyakov, Nucl. Phys. 164, 171 (1980).

[9] Chan Hong-Mo, Peter Scharbach and Tsou Sheung Tsun, Ann. Phys. (NY) 166, 396 (1986).

[10] I. Zois, hep-th/9703069, to appear in Rept. Math. Phys. (1998).

[11] E. Lubkin, Ann. Phys. (NY) 23, 233 (1963).

[12] S. Coleman, in *New Phenomena in Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1976), p. 297.

[13] P.A.M. Dirac, Proc. Roy. Soc. London A133, 60 (1931).

[14] C.N. Yang, Phys. Rev. D1, 2360 (1979).

[15] Chan Hong-Mo and Tsou Sheung Tsun, Nucl. Phys. B189, 364 (1981); see also [7].

[16] Chan Hong-Mo, Peter Scharbach and Tsou Sheung Tsun, Ann. Phys. (NY) 167, 454 (1986).

[17] Tai Tsun Wu and Chen Ning Yang, Phys. Rev. D14, 437 (1976).

[18] Chan Hong-Mo, Jacqueline Faridani, and Tsou Sheung Tsun, Phys. Rev. D51, 7040 (1995).

[19] C.H. Gu and C.N. Yang, Sci. Sin. 28, 483 (1975).

[20] Chan Hong-Mo, Jacqueline Faridani, and Tsou Sheung Tsun, hep-th/9512173, Phys. Rev. D53, 7293 (1996).

[21] S.K. Wong, Nuovo Cimento A65, 689 (1970).

[22] Chan Hong-Mo, Jacqueline Faridani, and Tsou Sheung Tsun, Phys. Rev. D52, 6134 (1995).

[23] Chan Hong-Mo, Jacqueline Faridani, Jakov Pfaudler, and Tsou Sheung Tsun, hep-th/9603188, Phys. Rev. D55, 5015 (1997).

[24] G. 't Hooft, Nucl. Phys. B138, 1 (1978); Acta Physica Austriaca Suppl. XXII, 531 (1980).

[25] Chan Hong-Mo and Tsou Sheung Tsun, hep-th/9702117, Phys. Rev. D56, 3646 (1997).

[26] See e.g. F. Fucito, M. Martellini, and M. Zeni, Nucl. Phys. B496, 259 (1997).