Nuclear Physics B190[FS3] (1981) 455–478

TOPOLOGY OF THE GAUGE CONDITION AND NEW CONFINEMENT PHASES IN NON-ABELIAN GAUGE THEORIES*

G. 't HOOFT[1]

California Institute of Technology, Pasadena, California 91125, USA

Received 9 March 1981

The gauge-fixing constraint in a gauge field theory is crucial for understanding both short-distance and long-distance behavior of non-abelian gauge field theories. We define what we call "non-propagating" gauge conditions such as the unitary gauge and "approximately non-propagating" or renormalizable gauge conditions, and study their topological properties. By first fixing the non-abelian part of the gauge ambiguity we find that SU(N) gauge theories can be written in the form of abelian gauge theories with $N-1$ fold multiplicity enriched with magnetic monopoles with certain magnetic charge combinations. Their electric charges are governed by the instanton angle θ.

If θ is continuously varied from 0 to 2π and a confinement mode is assumed for some θ, then at least one phase-transition must occur. We speculate on the possibility of new phases: e.g., "oblique confinement," where $\theta \simeq \pi$, and explain some peculiar features of this mode. In principle there may be infinitely many such modes, all separated by phase transition boundaries.

1. Introduction

It is a long-standing problem how to devise a reliable method for computing physically observable quantities accurately in non-abelian gauge theories with strong interactions. One crucial step in solving this problem is to isolate the relevant dynamical variables at the critical distance scale (in quantum chromodynamics that is, of course, the hadronic mass scale). Approximation methods that are popular at present sometimes ignore some of these variables. In ordinary perturbation theory the compactness of the gauge group is not reflected, so that topologically non-trivial effects such as instantons are not seen. In the lattice gauge theories this compactness is observed, but instantons literally slip through the meshes. Theories based on instanton gasses in their turn do not admit magnetic vortices, which we know to be crucial for understanding confinement. We do not (yet?) have a remedy for this situation. We do propose to make a new start although so far our considerations are

* Work supported in part by the US Department of Energy under contract no. DE-AC-03-79ER0068 and by the Fairchild Foundation.

[1]On leave from the Institute for Theoretical Physics, University of Utrecht, P.O. Box 80.006, 3508 TA Utrecht, The Netherlands.

qualitative; quantitative calculations are not carried out in this paper. Certainly our more distant aim is to set up a precise calculational procedure along the lines which we will explain.

Our philosophy is that isolation of the physically relevant parameters is in principle extremely simple. All we have to do is to fix the gauge, but this gauge-fixing must be done in such a way that no spurious, unphysical waves can propagate in our system. That happens for instance in the Lorentz gauge

$$\partial_\mu A_\mu = 0. \tag{1.1}$$

Such a gauge condition requires the solution of Laplace-like differential equations. For infinitesimal fields A'_μ the gauge transformation $\Lambda(x)$ turning it into the Lorentz gauge must satisfy

$$\partial_\mu D_\mu \Lambda = \partial_\mu A'_\mu, \tag{1.2}$$

and the inverse of $\partial_\mu D_\mu$ is a non-local operator. Indeed, this gauge is accompanied by a negative metric ghost particle [1,2] which further obscures the physical modes by the time we try to solve the theory's equations non-perturbatively.

In sects. 2 and 3 we construct what we call "non-propagating" or "non-communicating" gauge conditions. An example is the familiar "unitary gauge" in Higgs-Kibble theories [2,3]. They render the theory non-manifestly renormalizable, and therefore, at a later stage one should go over to the more promising "approximately non-communicating" gauges, where all gauge artifacts that propagate, including the ghosts, have a finite mass. The danger, however, is that then one might introduce not only massive ghosts, but also more troublesome unphysical long-distance features: topological structures such as planes or strings that are unobservable: the phantom solitons.

The strictly local, non-communicating gauges induce singularities in space-time. We make the important assertion that these singularities have a physical meaning; not their precise location, which might vary if one goes from one non-communicating gauge to another, but just their mere existence. We claim that they are the remaining physical dynamical variables besides the other physical fields.

In the abelian Higgs theory our gauge singularities are string-like, and they are to be interpreted as magnetic vortex tubes which are well known to occur in this theory [4].

In a non-abelian theory the simplest thing to do is to fix first the "non-abelian part" of the gauge redundancy, reducing the gauge symmetry to that of the maximal abelian subgroup. Here we get singularities which are point-like in 3-space (or particle-like in 4-space). If the gauge group is $SU(N)$, then our abelian subgroup becomes $U(1)^{N-1}$, so we get an effective theory with $N-1$ different kinds of electric charges. Many fields will have various combinations of these charges. We will then observe that our point-like singularities also play a dynamical role in this system: they are magnetic monopoles with respect to $U(1)^{N-1}$.

At this stage, then, our particle spectrum (not necessarily their interactions) is to some extent symmetric under the dual transformation electric $\leftrightarrow$ magnetic. We still do not have the dynamics under control so we do not know what the masses and interactions are. But since Bose condensation among electrically charged particles is known to be possible (as it happens in all superconductors), we can now easily imagine that, instead, the magnetic charges Bose condensate. If that happens we have absolute quark confinement.

The system of electrically and magnetically charged particles in an abelian gauge field, which we obtain in the non-communicating gauge, may be considered as "transient particles." They are the spectrum of particles propagating in a length/mass scale intermediate between the microscopic gauge theory and the macroscopic system of hadrons.

We think that this way of understanding the confinement phenomenon is more precise than similar arguments given earlier [5,6], and holds the promise that qualitative calculation procedures may be deduced.

In two previous publications [6] it was concluded that in a non-abelian gauge theory with a not too large gauge group essentially three phases are possible: the Higgs mode, the confinement mode, and a mode with long-range interactions, presumably of Coulomb type. However, it was tacitly assumed that instantons play no significant role. This is not always true, but those conclusions remain unaltered if we put the instanton angle $\theta=0$. At other values of θ a richer spectrum of possibilities emerges. Indeed, in sect. 10 we argue that in principle an infinity of different phases is possible.

2. The gauge conditions in the abelian Higgs model

The long-distance structure of the abelian Higgs model (superconductor) is well known. It is therefore fruitful to explain our arguments first for this case.

Let the lagrangian be

$$\mathcal{L}=-\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu}-D_\mu\phi^*D_\mu\phi-\tfrac{1}{2}\left(\phi^*\phi-F^2\right)^2, \tag{2.1}$$

with

$$F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu,$$

$$D_\mu\phi=\partial_\mu\phi+iqA_\mu\phi. \tag{2.2}$$

Here, ϕ is a complex field. A non-propagating gauge condition is

$$\mathrm{Re}(\phi)=\rho>0, \qquad \mathrm{Im}(\phi)=0. \tag{2.3}$$

It leads to well-known unitary Feynman rules, which are not manifestly renormalizable. In this gauge all components of A_μ correspond to observable fields: the massive vector field. And ρ corresponds to an observable scalar excitation: the Higgs particle [7].

There is, however, another structure at the long-distance scale: the magnetic vortex line [4]. But this vortex can also readily be identified in this gauge. The gauge condition (2.3) becomes singular at those points in space-time where ϕ happens to be zero. Since these points must satisfy two independent constraints on ϕ,

$$\mathrm{Re}(\phi)=0, \qquad \mathrm{Im}(\phi)=0, \tag{2.4}$$

they will form a two-dimensional subspace of Minkowski space, or a string in 3-space.

The equations for the vector field A_μ will be singular on this string but of course this singularity is an artifact of our gauge. What is not an artifact, however, is the fact that string-like objects are present: they are additional dynamical variables apart from the heavy particles which we had already; their topological structure reflects the topological nature of the system as a whole.

Note that the gauge condition (2.3) is possible as soon as a charged scalar field is available, whether or not $\langle\phi\rangle\neq 0$. We then have the dynamical variables listed above. But of course, if we are in the Coulomb phase, $\langle\phi\rangle=0$, then the string-like "singularity" is ubiquitous in the vacuum, the string condenses to form long-range magnetic field lines and our gauge choice is then not very suitable anymore to describe the long-distance structure.

For calculational purposes the unitary gauge (2.3) is not always practical because of its singular nature at small distances. The theory is not easy to be renormalized in this gauge. We now would like to preserve its good features at the long-distance scale but find a smoother condition at the microscopic scale: something in between the unitary gauge and the Lorentz gauge. Such gauges exist: non-physical features do propagate, but with a large mass-parameter so that they do little harm to the large-scale spectrum of dynamical variables. They have been used often [8]:

$$\arg(\phi)+\kappa\partial_\mu A_\mu=0, \tag{2.5}$$

where κ is a constant that we may choose and $\arg(\phi)$ defined to lie between $-\pi$ and π. Equivalently, one may add a gauge-fixing term to the lagrangian, for instance:

$$\delta\mathcal{L}=2qF^2\kappa^{-1}\cos\big(\arg(\phi)+\kappa\partial_\mu A_\mu\big) \tag{2.6}$$

(we chose the cosine only to avoid the singularity at $\arg(\phi)=\pm\pi$). The Faddeev-Popov ghost [1,2] happens to decouple:

$$\mathcal{L}^{\mathrm{gh}}=-\partial_\mu\eta^*\partial_\mu\eta-\frac{q}{\kappa}\eta^*\eta. \tag{2.7}$$

In both cases we find, in addition to the physical particles, ghosts that propagate with a mass

$$m_{\mathrm{gh}} = (q/\kappa)^{1/2}. \tag{2.8}$$

It is only in this simple model that they happen to decouple and the gauge (2.6) has the further convenience that the various ghosts do not mix.

The limit $\kappa \to 0$ clearly reproduces the unitary, non-propagating gauge (2.3). But how do the singular strings emerge in this limit? The answer is: here they do not. The gauge (2.5) is the gauge where

$$\int \mathrm{d}^4x \left(q^{-1}(\arg\phi)^2 + \kappa A_\mu^2 \right) \tag{2.9}$$

has a local minimum. At a string singularity, A_μ would diverge as (distance)$^{-1}$, and in a plane orthogonal to the string

$$\int \mathrm{d}^2x A_\mu^2 \tag{2.10}$$

would diverge logarithmically. In the space of continuous, differentiable A_μ the integral (2.9) is finite and we would have to search for a minimum there first.

Now consider a 3-dimensional space. Suppose we have a field configuration with zeros for ϕ, forming a string (of course this string has no ends). The gauge (2.3) then shows a singularity there. What happens then in the gauge (2.5) when κ tends to zero? The answer is that the system forms a sheet, with the aforementioned string at its edge. The thickness of the sheet is determined by $\sqrt{\kappa}$. The derivative in (2.5) prevents any singularities from developing. At either side of the sheet we have

$$\arg\phi \to 0. \tag{2.11}$$

To solve the equations for the sheet we make the simplifying assumption that it is essentially flat compared to the distance scale set by $(q/\kappa)^{1/2}$, and in the region of space which it occupies we neglect variations in the physically observable fields. The problem is then essentially one-dimensional. In the Lorentz-gauge we would have $\arg\phi \simeq 0$ and $A_\mu \simeq 0$. In the gauge (2.5) we have

$$\arg\phi + 2\pi n = \Lambda,$$

$$A_\mu = -q^{-1}\partial_\mu\Lambda, \tag{2.12}$$

$$\Lambda - \kappa q^{-1}\partial^2\Lambda = 2\pi n. \tag{2.13}$$

Here the integer n is such that $|\arg\phi| \leq \pi$. The boundary condition is such that ϕ makes a full rotation over 2π. The solution is a nice soliton:

$$\arg\phi = \pi\,\mathrm{sgn}(x)\exp\left\{-(q/\kappa)^{1/2}|x|\right\}, \tag{2.14}$$

where x is a coordinate orthogonal to the sheet.

The conclusion of this section is that the transition from a unitary, non-propagating gauge to a renormalizable gauge produces an important change in the set of dynamical variables. The unitary gauge shows clearly the physical objects: heavy vector particles, scalar particles and also strings. The renormalizable gauge produces not only ghosts but also phantom sheets. If $\kappa \to 0$ then the ghosts become infinitely heavy and the sheets infinitely thin.

How could we re-obtain the most relevant part of the dynamical variables also in the renormalizable gauge? Suppose that we reintroduce the string singularities explicitly. Along those singularities the integral (2.9) would diverge, but we can in principle invent a subtraction procedure, recovering a finite part. Then we can ask for that particular configuration of the singularities that minimizes this finite part of the integral (2.9). The soliton (2.14) contributes a lot:

$$2\pi^2\kappa^{1/2}q^{-3/2}\int \mathrm{d}^2x, \tag{2.15}$$

which diverges with the area. So large sheets will then automatically be avoided and be exchanged for strings. Thus also in renormalizable gauges the correct set of dynamical variables can be obtained.

A quantitative treatment of the theory would now require a study of the continuum theory in the renormalizable gauge with some isolated string-like singularities. We will not pursue such programs any further but it is clear that in principle all long-distance phenomena would be taken in account correctly. The more important point we wish to make is that the same can be done for non-abelian gauge theories and that such a program again assures us of the correct dynamical long-distance variables.

In the non-abelian case however it is best to follow such a procedure in two steps. First we fix the non-abelian part of the gauge. The dynamical variables thus obtained are the "transient variables." Then afterwards one can treat the abelian part, which is much more conventional.

As a matter of fact we will show that during the first part of this procedure, which is the most crucial one, we can avoid the bothersome string singularities. The real singularities, then, are just point-like and won't be wiped out in the more regular renormalizable gauges.

3. The abelian projection in the non-abelian gauge theory

In the non-abelian theory, whether or not spontaneously broken, we again search first for a non-propagating gauge condition. Clearly, the Lorentz gauge

$$\partial_\mu A_\mu = 0, \tag{3.1}$$

propagates massless spurious modes and is therefore unsuitable. One must use some tensor, X, that transforms covariantly under a gauge transformation Ω:

$$X' = \Omega X \Omega^{-1}. \tag{3.2}$$

Since scalar fields or field combinations transforming as a fundamental representation of the gauge group are often not present we concentrate on the case that X is in the adjoint representation. For instance

$$X = G_{\mu\nu} G_{\mu\nu}, \tag{3.3a}$$

$$X = G_{\mu\nu} D^2 G_{\mu\nu}, \tag{3.3b}$$

or

$$X = G_{12}, \tag{3.3c}$$

where $G_{\mu\nu}$ is the covariant curl in matrix notation:

$$G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - gi\left[A_\mu, A_\nu\right]. \tag{3.4}$$

Choice (3.3a) only makes sense if the gauge group is larger than SU(2), and there are perhaps other possibilities as we shall see.

How does X determine the gauge? The eigenvalues of X are gauge invariant. Therefore, the best we can do is search the gauge in which X is diagonal:

$$X = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix}. \tag{3.5}$$

Just as in the abelian model, this gauge condition produces singularities. Here, the singularities occur when two or more eigenvalues coincide. The nature of these singularities now depends on the gauge group.

If the gauge group is an orthogonal group SO(N) then at non-singular points (3.5) fixes the gauge completely, and the singular points, with two coinciding eigenvalues, form strings in 3-space. These represent Nielsen-Olesen vortices [4] and the remainder of the discussion would be very similar to that of sect. 2. Only, if no spontaneous breakdown takes place,

$$\langle X\rangle = \lambda I, \tag{3.6}$$

then these strings will fill the vacuum, and further discussion of that situation will be hard. It is precisely these string-like singularities which we would like to avoid, since field theories for strings are hard and unconventional. We shall not elaborate further on the question how to find a non-propagating gauge condition with only point-like singularities for SO(N).

If the gauge group is SU(N) the gauge (3.5) is particularly interesting. At generic points where the λ_i do not coincide, the gauge is not determined completely, since any diagonal gauge rotation Ω,

$$\Omega = \begin{pmatrix} e^{i\phi_1} & & \\ & \ddots & \\ & & e^{i\phi_N} \end{pmatrix}, \qquad \Sigma\phi_i = 0, \tag{3.7}$$

leaves X invariant if X looks like (3.5) and transforms according to (3.2). In fact, the subgroup of Ω satisfying (3.7) is the largest abelian subgroup,

$$\mathrm{U}(1)^{N-1}, \tag{3.8}$$

so within this gauge our system is an $N-1$ fold abelian gauge-invariant theory. We will call this group for simplicity: the "electromagnetic" group. The diagonal components of a matrix are "neutral," and off-diagonal components of a matrix carry two different "electric" charges.

However, our system has something else besides electromagnetic gauge fields and electric charges. In this gauge there are also singularities, namely if two eigenvalues λ of X coincide. This time it is easy to establish that such exceptional points occur only if three conditions are met, and so they form isolated points in 3-space, not strings. For instance in SU(2) the conditions that the hermitian matrix

$$X = a_0 + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 \tag{3.9}$$

(where σ_k are the Pauli matrices) has two coinciding eigenvalues are

$$a_1 = a_2 = a_3 = 0; \tag{3.10}$$

clearly three conditions, fixing the three space coordinates of such points. But also for all other values of N the condition that two eigenvalues of an hermitian $N\times N$ matrix coincide forms three constraints. The singularities, therefore, are always point-like. It appears that gauge conditions with only point-like singularities are only

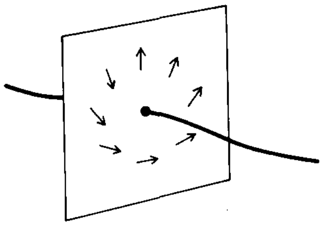

Fig. 1. In a plane orthogonal to the string singularity in gauge (3.12) the axis of the invariant U(1) rotation in isospin space can rotate as indicated here.

possible if the largest abelian subgroup of the gauge group is left unbroken. We call this the abelian projection.

Note that if the gauge group is SU(2) then the choice (3.3a) for X is not allowed: $G_{\mu\nu}G_{\mu\nu}$ automatically has coinciding eigenvalues. If we use vector notation,

$$G_{\mu\nu}=\tfrac{1}{2}\tau^a G^a_{\mu\nu}, \tag{3.11}$$

where τ^a are the Pauli matrices, then one could be tempted to use

$$G^0_{\mu\nu}G^1_{\mu\nu}=G^0_{\mu\nu}G^2_{\mu\nu}=0 \tag{3.12}$$

as a gauge condition that leaves a U(1) invariance. However, $G^a_{\mu\nu}G^b_{\mu\nu}$ is a real symmetric matrix, and the locus of points with two coinciding eigenvalues is string-like. In the vicinity of such a string the axis of the invariant U(1) group may rotate as pictured in fig. 1. Because of this disease we prefer an X transforming as (3.2), and of the form (3.3b), or (3.3c) if the gauge group is SU(2).

4. The nature of the singularities

The gauge in which X is diagonal, see e.g. (3.5), can be further restricted by choosing

$$\lambda_1>\lambda_2\cdots>\lambda_N. \tag{4.1}$$

The only singularities occur if two consecutive eigenvalues coincide:

$$\lambda_j\Rightarrow\lambda_{j+1}\equiv\lambda, \qquad \text{for certain } j. \tag{4.2}$$

Let us consider the direct vicinity of such a point. We ignore all other rows and columns of X except the jth and $j+1$st. Prior to complete diagonalization the matrix was

$$X=\left(\begin{array}{c|cc|c} D_1 & \multicolumn{2}{c|}{0} & 0 \\ \hline 0 & \lambda+\varepsilon_3 & \varepsilon_1-i\varepsilon_2 & 0 \\ & \varepsilon_1+i\varepsilon_2 & \lambda-\varepsilon_3 & \\ \hline 0 & \multicolumn{2}{c|}{0} & D_2 \end{array}\right)\equiv\lambda+\varepsilon_k\sigma_k, \tag{4.3}$$

using a short-hand notation. Since all other eigenvalues differ from λ we could safely take D_1 and D_2 to be diagonal. ε_k are small. With respect to that SU(2) subgroup of our gauge group that involves only the jth and $j+1$st components of a fundamental representation, the fields $\varepsilon_k(x)$ behave in all respects as an isovector.

Our singularity occurs at x_0, where

$$\boldsymbol{\varepsilon}(x_0)=0, \tag{4.4}$$

and our gauge condition corresponds to rotating $\boldsymbol{\varepsilon}$ such that

$$\varepsilon_3>0, \qquad \varepsilon_1=\varepsilon_2=0, \tag{4.5}$$

leaving invariance with respect to rotations around the 3-axis. It is now well known that in this gauge the zero-point of $\boldsymbol{\varepsilon}$ at x_0 behaves as a magnetic charge with respect to the remaining U(1) rotations [9]. This magnetic charge h is such that if the U(1) charges of a doublet

$$\begin{pmatrix} \psi_j \\ \psi_{j+1} \end{pmatrix}$$

are defined to be $\pm\frac{1}{2}g$, then

$$hg=4\pi. \tag{4.6}$$

For labeling these charges it is convenient momentarily to replace SU(N) by U(N), whose largest abelian subgroup is

$$\mathrm{U}(1)^N, \tag{4.7}$$

being the group of rotations

$$\Omega=\begin{pmatrix} e^{i\omega_1} & & \\ & \ddots & \\ & & e^{i\omega_N} \end{pmatrix}. \tag{4.8}$$

The U(1) invariant subgroup of rotations $\omega_1 = \omega_2 = \cdots = \omega_N \equiv \omega$ can then easily be split off afterwards.

Let us then label magnetic charges as a vector $(h_1 \dots, h_N)$. We choose units such that the ith component of a fundamental representation $(\psi_1, \dots, \psi_i, \dots, \psi_N)$ has electric charge $(0, \dots, 0, \frac{1}{2}g, 0, \dots, 0)$ with the $\frac{1}{2}g$ at the ith position. Then our singularity has magnetic charge

$$h_j = -h_{j+1} = 4\pi/g,$$
$$h_i = 0, \qquad \text{for } j \neq i \neq j+1. \tag{4.9}$$

Thus we see that our singularities come in $N-1$ different varieties, all carrying two consecutive and opposite magnetic charges. "Monopoles" with non-consecutive opposite charges can only be obtained as "bound states" of these fundamental monopole singularities.

We conclude in this section that the non-abelian SU(N) gauge theory is topologically such that it can be cast into a $\mathrm{U}(1)^{N-1}$ abelian gauge theory, which, however, will feature not only electrically charged particles but also magnetic monopoles with magnetic charges given by (4.9). However, the gauge used in this mapping renders the theory non-manifestly renormalizable. In the next section we introduce a smoother gauge condition.

4. The approximately non-propagating gauge

It is easy to write down a renormalizable gauge condition that reduces SU(N) to $\mathrm{U}(1)^{N-1}$. If we represent the vector fields by hermitian matrices $A_\mu(x)$ then we may define

$$A_\mu = A_\mu^0 + A_\mu^{\mathrm{ch}}. \tag{5.1}$$

Here A_μ^{ch} is the matrix A_μ with all diagonal elements replaced by zero. The suffix ch stands for "charged" with respect to the abelian $\mathrm{U}(1)^{N-1}$ rotations. We write A_μ^0 for the diagonal elements of A_μ only. A renormalizable gauge condition would be

$$D_\mu^0 A_\mu^{\mathrm{ch}} = 0, \tag{5.2}$$

where

$$D_\mu^0 = \partial_\mu - ig\left[A_\mu^0, \ \right]. \tag{5.3}$$

One easily verifies that this leaves the subgroup $\mathrm{U}(1)^{N-1}$ invariant. But, as in sect. 1, we argue that in this gauge massless ghosts propagate, which we wish to avoid.

An intermediate gauge could be the non-abelian equivalent of (2.5):

$$X^{\mathrm{ch}} - \kappa i\left[D_\mu^0 A_\mu^{\mathrm{ch}}, X^0\right] = 0, \tag{5.4}$$

which corresponds to (5.2) at small distances and (3.5) at large distance scales. If we take a patch of space-time where all eigenvalues of X differ from each other, then all propagators there have non-physical parts at a mass value

$$m = \mathrm{O}(g/\kappa)^{1/2}, \tag{5.5}$$

which is also the mass of the Faddeev-Popov ghosts. g is the coupling constant. By adjusting κ we can choose this mass to be anything we like. The special choice of eq. (5.4) ensures that this mass does not depend on the vacuum values of X.

Unfortunately, even at small distances (or equivalently at large κ) the gauge condition (5.4) is hard to implement. Perturbation expansion must be refined considerably. Further study of this will be necessary, and perhaps one has to choose even more refined gauge-fixing procedures. But let us study this one a bit more first.

A slightly modified version of (5.4) can be obtained from an extremum principle. Let us define a hermitian matrix Y having only non-diagonal elements, and

$$[Y, X^0] = iX^{\mathrm{ch}}. \tag{5.6}$$

In matrix notation, if the diagonal elements of X^0 are given by λ_i, then

$$\begin{aligned} Y_{ij} &= iX_{ij}(\lambda_j - \lambda_i)^{-1}, &\quad &\text{if } i \neq j, \\ &= 0, & &\text{if } i = j. \end{aligned} \tag{5.7}$$

Under an infinitesimal gauge transformation given by a matrix Λ^{ch} (while $\Lambda^0 = 0$), this transforms as

$$Y' = Y - \Lambda^{\mathrm{ch}} + \mathrm{O}(\Lambda^{\mathrm{ch}}, Y), \tag{5.8}$$

where, as always, the suffix ch stands for removing the diagonal components. If we replace (5.4) by

$$Y + \kappa D_\mu^0 A_\mu^{\mathrm{ch}} + \mathrm{O}(Y^2) = 0, \tag{5.9}$$

then that is obtained by minimizing

$$W = \int \mathrm{d}^4x \,\mathrm{Tr}\left(g^{-1} Y^2 + \kappa\left(A_\mu^{\mathrm{ch}}\right)^2\right), \tag{5.10}$$

under gauge transformations.

6. Phantom strings or singularities?

For any non-vanishing value of κ one would not expect monopole-like singularities such as the ones discussed in sect. 4. However, the integral (5.10) has to be finite. At a monopole singularity itself, Y and A_μ^{ch} diverge only as $1/r$, so that there the integral does converge. Therefore, the system might be unstable against the formation of singularities, as we will see.

Let us assume that two adjacent eigenvalues, λ_j and λ_{j+1}, coincide at two points A and B, and that the non-propagating gauge (3.5) would produce a monopole at A and one with opposite magnetic charges at B. If these singularities are to be avoided in the more regular gauge (5.9) then X^{ch} cannot be chosen to vanish everywhere. The configuration with minimal value for W [eq. (5.10)] would therefore show a structure extending from A to B with non-vanishing Y and A_μ^{ch}. This structure, which would clearly be a gauge artifact, will be called "phantom string".

However, it so happens that here the string-like phantom soliton is unstable against collapse. One reason is that in our gauge, the integral W, eq. (5.10), gets an infinite contribution where two diagonal elements (not two eigenvalues) of X coincide. Another is that the field variable here is just the gauge transformation Ω, a scalar field, and such solitons are usually unstable in two dimensions. This fact constitutes a practical, not fundamental, difference with the abelian case, as described in sect. 2, where the new phantom was stable. It just happens to be true, and it is very convenient. The approximately non-propagating gauge (5.9) will automatically produce monopole-like singularities. And, furthermore, outside those singularities two diagonal elements of X are never allowed to be equal. This enables us to order them:

$$\lambda_1 > \lambda_2 \cdots > \lambda_N, \tag{6.1}$$

everywhere except at the singularities.

7. A renormalizable gauge: phantom surfaces

Even though the gauge (5.4) tends to the Lorentz gauge at small distances, one may expect that it is tedious to be implemented in an ordinary perturbation expansion for the small distance phenomena. The reason is that the quantity Y in (5.8) is singular. A gauge condition which is guaranteed to be safe in this respect is the one obtained by extremizing not the function W of eq. (5.10), but instead

$$W_2 = \int \mathrm{d}^4x \operatorname{Tr}\Big(g^{-1}(X^{\text{ch}})^2 + \kappa\big(A_\mu^{\text{ch}}\big)^2\Big). \tag{7.1}$$

One then obtains the gauge condition:

$$[X^{\text{ch}}, X^0] - \kappa i D_\mu^0 A_\mu^{\text{ch}} = 0. \tag{7.2}$$

The reason why we did not take this gauge from the start is that now the ghost mass is much less predictable. In perturbation expansion it is zero for all κ. In any background with non-vanishing X the masses of ghosts with charges i and j are

$$m_{\mathrm{gh}} = \mathrm{O}\left[(g/\kappa)^{1/2}|\lambda_i - \lambda_j|\right], \tag{7.3}$$

where λ_i are the eigenvalues of X. These ghosts therefore look much more troublesome but still, in the limit $\kappa \to 0$ the non-propagating gauge should be reached.

Again, this gauge is unstable against the formation of monopole singularities, so that phantom string-like solitons do not emerge, but another phantom can be produced in principle: phantom sheets. This is because now the eigenvalues λ_i of X^0 cannot be kept ordered. We get various regions of space-time where their order will be different. These regions will be separated by domain walls, much like Bloch walls in a ferromagnet. Let us take a closer look at these walls.

Again we take X to be in the representation (4.3), neglecting all but two rows and columns:

$$X = \lambda + a_k \sigma_k. \tag{7.3}$$

We are far from any singular points, so

$$|\boldsymbol{a}| = a \neq 0. \tag{7.4}$$

We take κ to be small with respect to the scale in which the physical fields vary. Then we are close to a pure gauge transformation of the approximately constant field configuration

$$X'_\mu = \lambda + a'_3\sigma_3, \qquad A'_\mu = 0. \tag{7.5}$$

This configuration is transformed by a gauge transformation $\Omega(x)$ such that at $x_3 \gg 0$ we have

$$X \equiv \Omega X' \Omega^{-1} = \lambda + a'_3\sigma_3,$$
$$A_\mu \equiv i g^{-1} \Omega \partial_\mu \Omega^{-1} = 0, \tag{7.6}$$

and at $x_3 \ll 0$ we have

$$X = \lambda - a'_3\sigma_3;\ A_\mu = 0. \tag{7.7}$$

Ω must be of the form

$$\Omega = \exp(i\omega\sigma_1),$$

where ω depends on x_3, and

$$W_2 = \int \mathrm{d}^4x \left\{ g^{-1}\left((\Omega X' \Omega^{-1})^{\mathrm{ch}} \right)^2 - g^{-2}\kappa \left(\left(\Omega \partial_\mu \Omega^{-1} \right)^{\mathrm{ch}} \right)^2 \right\}$$

$$= 2 \int \mathrm{d}^4x \left\{ a^2 g^{-1} \sin^2 2\omega + g^{-2}\kappa \left(\partial_\mu \omega \right)^2 \right\}. \tag{7.8}$$

This is the lagrangian for a sine-Gordon equation, and the boundary conditions (7.6) and (7.7) are those of the sine-Gordon soliton. The contribution of the sheet to W_2 is:

$$W_2 = 4a\kappa^{1/2} g^{-3/2} \int \mathrm{d}^2x, \tag{7.9}$$

where the integral is over the sheet. The thickness of the sheet is of order

$$a^{-1}\kappa^{1/2} g^{-1/2} = \mathrm{O}\left(m_{\mathrm{gh}}^{-1} \right). \tag{7.10}$$

We may safely assume that the phantom sheets contribute to W_2 by an amount proportional to their area. Since they have no natural boundary they only come in the form of bubbles, and minimizing W_2 will correspond to minimizing the total area of these bubbles, which will therefore never grow to substantial sizes. So, in the bulk of space-time, with only small exceptional regions, we may assume that the λ_i are ordered in the same way as in the previous gauges, see (4.1) and (6.1). Monopole singularities will occur in this gauge as much as in the previous gauges. Because of the above their magnetic charges will be consecutive [in the sense of sect. 4, eq. (4.9)] in the "regular" regions of space-time, but non-consecutive charges might show up occasionally, inside a bubble. We see that although gauges of the type (7.2) might be easier to handle in perturbation theory, they give rise to a more complicated topology.

We conclude from sects. 3–7 that perhaps the most crucial physical dynamical variables that govern the strong-interaction region of a non-abelian theory may be obtained by first fixing the non-abelian part of the gauge symmetry. We get an abelian theory with magnetic monopoles. Their magnetic charges [in the case of SU(N)] are given by eq. (4.9). The short-range, non-electromagnetic interactions between the electrically charged particles and the monopoles are likely to be computable by ordinary perturbative techniques. If so, then perhaps one may be able to determine which of the corresponding fields develops non-vanishing vacuum expectation values, and thereby prove or disprove the phenomenon of confinement in various models. The remaining sections are devoted to the role instantons play in our set of transient dynamical variables.

8. The θ vacuum

Dirac [10] observed that the equations of motion of an electrically charged particle with charge q in the vicinity of a magnetic monopole field with source strength h can only be quantized in a rotationally invariant way if

$$hq = 2\pi n, \tag{8.1}$$

where n is an integer.

Now we are dealing with a multiple abelian theory, with gauge group $U(1)^N$. Furthermore, our monopoles might also carry electric charge. Let us consider then two different kinds of particles, 1 and 2, with magnetic charges $h_i^{(1)}$ and $h_i^{(2)}$, and electric charges $q_i^{(1)}$ and $q_i^{(2)}$, where the index i refers to the particular U(1) group. Dirac's condition (8.1) corresponds to quantizing the Lorentz force acting between these particles, so in the more general case it reads

$$\sum_{i=1}^{N} \left(h_i^{(1)} q_i^{(2)} - q_i^{(1)} h_i^{(2)} \right) = 2\pi n. \tag{8.2}$$

We will sometimes refer to this equation in words: "the particles (1) and (2) have a Dirac unit n with respect to each other." Since the overall U(1) rotations are of no concern to us we make the restriction

$$\sum_i h_i = \sum_i q_i = 0. \tag{8.3}$$

Clearly this restriction merely determines a notational convention, useful for describing the $U(1)^{N-1}$ abelian subgroup of $SU(N)$.

If two particles satisfy (8.2) and (8.3) then a third, with charges

$$h_i^{(3)} = k_1 h_i^{(1)} + k_2 h_i^{(2)},$$

$$q_i^{(3)} = k_1 q_i^{(1)} + k_2 q_i^{(2)}, \tag{8.4}$$

where k_1 and k_2 are integers, satisfies the correct quantization conditions with respect to both others. Therefore, the most general spectrum of particles satisfying (8.2) and (8.3) is obtained by first finding a basis of $2(N-1)$ particles with charges $h_i^{(A)}$ and $q_i^{(A)}$; $A = 1,\dots,2N-2$. All allowed sets of charges are then

$$h_i = \sum_{A=1}^{2N-2} k_A h_i^{(A)}, \qquad q_i = \sum_{A=1}^{2N-2} k_A q_i^{(A)}. \tag{8.5}$$

They form a $2N-2$ dimensional lattice.

Maxwell's equations are invariant under the rotations

$$h_i \to h_i \cos\phi_i + q_i \sin\phi_i,$$
$$q_i \to -h_i \sin\phi_i + q_i \cos\phi_i. \tag{8.6}$$

Therefore, for $N-1$ of the $2N-2$ basis elements we can rotate away the magnetic charge:

$$h_i^{(A)} = 0, \qquad \text{for } A = 1,\dots,N-1. \tag{8.7}$$

For the electrically charged gluons in our system we find a basis

$$q_i^{(A)} = \tfrac{1}{2} g \delta_i^A - \tfrac{1}{2} g \delta_i^{A+1}, \qquad \text{for } A = 1,\dots,N-1. \tag{8.8}$$

The magnetic monopoles have magnetic charges according to eq. (4.9):

$$h_i^{(A)} = \frac{4\pi}{g} \delta_i^{A+1-N} - \frac{4\pi}{g} \delta_i^{A+2-N}, \qquad \text{for } A = N,\dots,2N-2.$$

Note that "quarks" may be considered to have only electric charges:

$$q_i = \tfrac{1}{2} g \delta_{ii_0} - \frac{g}{2N}, \tag{8.10}$$

where the last term is an overall U(1) charge, added in order to comply with (8.3). We see immediately that N quarks, or a quark and an antiquark are needed to obtain something that fits in the lattice spanned by the gluons. Also, the quarks would saturate Dirac's condition, the gluons alone do not.

Coming back to the magnetically charged particles, we have not yet specified their electric charges: $q_i^{(A)}$ for $A = N,\dots,2N-2$. Any set of values would be consistent with Dirac's condition (8.2). It was Witten [11] who first observed that monopoles may indeed carry fractional electric charges when generated by a non-abelian gauge theory:

$$q_i^{(A)} = \frac{\theta g^2}{16\pi^2} h_i^{(A)}, \qquad A = N,\dots,2N-2, \tag{8.11}$$

where θ is the instanton-angle.

A simplified explanation of eq. (8.11), in a notation adapted to our $U(1)^{N-1}$ theory, goes as follows. If we decide not to tamper with the boundary conditions then the lagrangian in euclidean space for a θ vacuum is

$$\mathcal{L} = -\tfrac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} + \frac{\theta i g^2}{32\pi^2}G^a_{\mu\nu}\tilde{G}^a_{\mu\nu} + A^a_\mu J^a_\mu$$

$$= -\tfrac{1}{2}\mathrm{Tr}\,G_{\mu\nu}G_{\mu\nu} + \frac{\theta i g^2}{16\pi^2}\mathrm{Tr}\,G_{\mu\nu}\tilde{G}_{\mu\nu} + \mathrm{Tr}\,A_\mu J_\mu, \tag{8.12}$$

where $\tilde{G}_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}G_{\alpha\beta}$. The last term is an auxiliary source term that vanishes in the pure gauge theory. We assume that at some distance from the monopole singularity only the abelian parts survive:

$$\mathcal{L} = \sum_i \left(-\tfrac{1}{2}F_{\mu\nu i}F_{\mu\nu i} + \frac{\theta i g^2}{16\pi^2}F_{\mu\nu}\tilde{F}_{\mu\nu i} + A_{\mu i}J_{\mu i} \right), \tag{8.13}$$

where the index i refers to the ith U(1) group. Some rather unconventional factors of 2 are due to our transition towards the diagonal part of the group. In the gauge

$$A_{4i} = 0, \tag{8.14}$$

this reads

$$\mathcal{L} = \sum_i \left(-(\partial_\tau \boldsymbol{A}_i)^2 - \boldsymbol{b}_i^2 + \frac{\theta i g^2}{4\pi^2}\partial_\tau \boldsymbol{A}_i \cdot \boldsymbol{b}_i + \boldsymbol{A}_i \cdot \boldsymbol{J}_i \right), \tag{8.15}$$

where τ is euclidean time and if $\boldsymbol{B}_i = \mathrm{curl}\,\boldsymbol{A}_i$, then $\boldsymbol{b}_i$ is $\boldsymbol{B}_i$ with its string singularity removed. For the transition towards Minkowski space we replace $\partial_\tau \boldsymbol{A}$ by $i\partial_t \boldsymbol{A} = i\dot{\boldsymbol{A}}$, so that

$$\mathcal{L} = \sum \left(\dot{\boldsymbol{A}}_i^2 - \boldsymbol{b}_i^2 - \frac{\theta g^2}{4\pi^2}\dot{\boldsymbol{A}}_i \cdot \boldsymbol{b}_i + \boldsymbol{A}_i \cdot \boldsymbol{J}_i \right). \tag{8.16}$$

The Euler-Lagrange equations are

$$\boldsymbol{D}_i = \dot{\boldsymbol{A}}_i - \frac{\theta g^2}{8\pi^2}\boldsymbol{b}_i, \tag{8.17a}$$

$$\dot{\boldsymbol{D}}_i = -\mathrm{curl}\left(\boldsymbol{b}_i + \frac{\theta g^2}{8\pi^2}\dot{\boldsymbol{A}}_i \right) + \tfrac{1}{2}\boldsymbol{J}_i. \tag{8.17b}$$

Therefore

$$\mathrm{div}\dot{\boldsymbol{D}}_i = \tfrac{1}{2}\mathrm{div}\boldsymbol{J}_i = -\tfrac{1}{2}\dot{\rho}_i. \tag{8.18}$$

The conserved quantity $\text{div}\boldsymbol{D}_i + \frac{1}{2}\rho_i$ must be chosen to be zero for the physical sector of Hilbert space. If no electric charge is present, then

$$\text{div}\boldsymbol{D}_i = 0 \tag{8.19}$$

The hamiltonian is

$$\mathcal{H} = \sum_i \left(\left(\boldsymbol{D}_i + \frac{\theta}{8\pi^2} g^2 \boldsymbol{b}_i \right)^2 + \boldsymbol{b}_i^2 - \boldsymbol{J}_i \cdot \boldsymbol{A}_i \right). \tag{8.20}$$

The electric field, as would be measured experimentally, is

$$\boldsymbol{E}_i = \tfrac{1}{2} \left(\boldsymbol{D}_i + \frac{\theta}{8\pi^2} g^2 \boldsymbol{b}_i \right) = \tfrac{1}{2} \dot{\boldsymbol{A}}_i. \tag{8.21}$$

Pure magnetic monopoles must satisfy (8.19), and due to (8.17a) the field $\boldsymbol{D}_i$ has no string singularity for stationary monopoles. Therefore, for stationary monopoles $\boldsymbol{D}_i$ vanishes, and they carry along with them an electric field

$$\boldsymbol{E}_i = \frac{\theta g^2}{16\pi^2} \boldsymbol{b}_i, \tag{8.22}$$

corresponding to an electric charge as given in (8.11). This result appears to be non-periodic in θ but that is not so. The result at $\theta = 2\pi$ may be seen as a bound state of a monopole and a gluon. It will be important to note that the monopole and the gluon in this bound state have a relative Dirac quantum of two units, regardless the value of N. If this quantum would have been odd then this bound state of two bosons would have been a fermion [12] and we could not possibly have exact periodicity in θ with period 2π.

The phenomenon of a metamorphosis of a pure magnetic flux into a magnetic flux plus an electric flux when θ runs from 0 to 2π has been derived in a different setting in ref. [13].

Eqs. (8.7)–(8.9) and (8.11) define the basis of our lattice of existing electromagnetic charge combinations.

9. A phase transition in θ

From the previous sections we obtain the following picture of the transient dynamical variables for a non-abelian gauge theory: there will be a multiple set of abelian Maxwell fields (if the original gauge group was SU(N), their multiplicity is $N-1$, to be indicated with an index $i = 1, \ldots, N$ with a restriction on the sum of all charges). There will be a number of electrically charged and magnetically charged particles, whose charges can be plotted as a charge-lattice of $2N-2$ dimensions. The SU(2) lattice is sketched in fig. 2. Apart from the obvious abelian interaction these

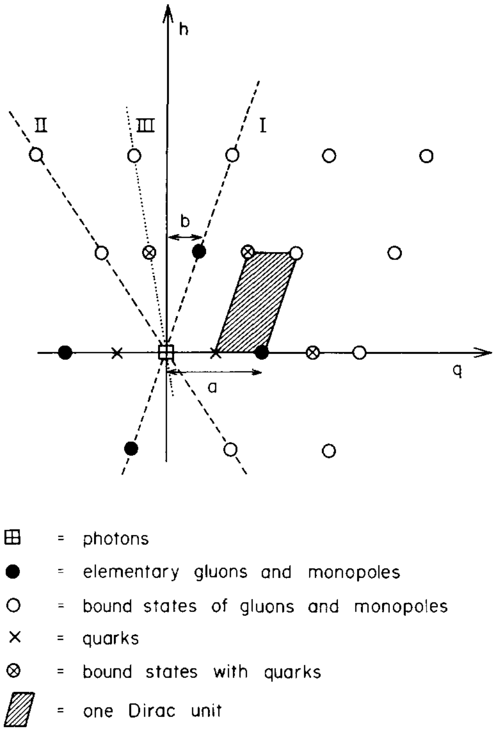

Fig. 2. The electric-magnetic charge lattice for the SU(2) case. q = electric charge, h = magnetic charge. $b/a=\theta/2\pi$. Dotted lines I, II, and III indicate Bose condensation in various phases.

particles may exert short-range non-gauge interactions due to particle exchange at small distances. The choice of the axes in this lattice is to some extent arbitrary because the Maxwell equations are invariant under orthogonal rotations in the h-q plane.

For the behavior of the theory at long-distance scales there are now three natural possibilities already often discussed in the literature: if one of the purely electrically charged objects is a Lorentz scalar it can develop a non-vanishing vacuum expectation value and we have the "Higgs mode." If the field of a magnetic monopole develops a non-vanishing vacuum expectation value we have permanent quark confinement ("confinement mode"). If all particles in the lattice get an ordinary positive mass-squared then we keep the physical particles as indicated in the lattice ("Coulomb mode"). The system cannot continuously enter one such mode from another. There must be sharp phase transition boundaries.

All phases can now be characterized by designating those points on the charge lattice that develop vacuum expectation values. The relative Dirac quantum for all pairs of these points must vanish. For the SU(2) case (fig. 2) that implies that they all must lie on a straight line through the origin. In the general case they can span at most an $N-1$ dimensional linear subspace, F. All particles whose charges lie in this subspace only show short-range interactions. Their gauge fields are screened by the Higgs mechanism. All particles that have a non-vanishing Dirac unit with respect to one of the points on this subspace F behave like monopoles in a superconductor: they are endpoints of a string that binds them to opposite charges.

Clearly then, the phase of the system is characterized by this linear subspace spanned by at most $N-1$ points on the lattice. (The pure Coulomb mode corresponds to only choosing the origin). Phase transitions correspond to discrete jumps replacing one linear subspace by another.

Suppose we had a confinement mode, corresponding to the dotted line I in fig. 2. If θ runs from 0 to π this dotted line becomes more tilted. At $\theta=\pi$ the other line, line II, corresponds to the parity image of I, and at $\theta \to 2\pi$ that new space, line II, merges again with the vertical axis. Clearly a phase transition is needed at least for one value of θ ($\theta=\pi$), or several values. This was derived by different but equivalent methods in ref. [14].

10. Oblique confinement

If $\theta \simeq \pi$ then neither I nor II (see fig. 2) are good candidates for the direction in which a vacuum expectation value develops. The monopole particles that have to move collectively then carry large electric charges. But the monopoles corresponding to I and II carry opposite electric charges. Perhaps they form a tight bound state (III) which in turn "condenses" (develops a vacuum expectation value). That would be a fourth mode, to be referred to as "oblique confinement." We will now argue that quarks in this mode will not be confined in the usual sense.

A quark or antiquark (indicated by crosses in fig. 2) may pick up at not much cost a monopole to form a bound state (cross in circle). One of these bound states is on line III and can therefore escape as a free particle, carrying the quark's flavor quantum numbers. However, something else happens that is worth one's attention. According to a strict and well known rule of quantum mechanics, the bound states of two particles is a boson or fermion depending on whether the constituents were bosons or fermions. What is not so well known, however, is that if two particles have an odd relative Dirac quantum then the rule is opposite from the usual one [12]: two bosons make a fermion; a boson and a fermion make a boson, etc. The orbital motion will show half-odd integer spin, so the spin-statistics theorem remains valid. Thus, if the quark was a fermion and the monopole a boson, then the freely moving particle ($\otimes$ on line III in fig. 2) is a boson: the quark escapes, but had to flip its spin and statistics properties.

This observation enables one to construct an unusual model with fermionic gauge particles but without supersymmetry. We start with SU(3) and instanton angle $\theta \simeq \pi$. There are no fermions. Assume that a fundamental scalar triplet field develops a vacuum expectation value, bringing the local symmetry down to SU(2). Assume that this SU(2) condenses in an oblique confinement mode. Four of the five heavy SU(3) gauge bosons form a complex SU(2) doublet. They are not confined, but escape disguised as fermions.

We do not know if it is possible to make light or massless fermions this way. This would require a chiral symmetry of which the original model showed no trace. Of course, chiral fermions in the original theory would wipe out any θ dependence. We argued in a previous publication that the discontinuity in θ must then probably be replaced by a spontaneous breakdown of the chiral symmetry among the fermions [14].

Of course oblique confinement can also be visualized in gauge theories with larger gauge groups. In SU(3) an oblique confinement mode that liberates the quarks occurs if a bound state of three monopoles and a gluon condenses. Table 1 lists the magnetic and electric charges of the various building blocks, and those of the condensing states. We see that if $\theta=\frac{2}{3}\pi$ then $3S=Q$ and the electric charges of the condensed particles vanishes so then this mode might occur. The last blocks in the table show which quark-monopole bound states become liberated, having charges that are linear combinations of those of the condensed states.

When Dirac's quantum for these bound states is worked out we find it to be even. Therefore the SU(3) quarks do not change from fermions into bosons or vice versa when liberated.

TABLE 1
Oblique confinement in SU(3)

		h_1	h_2	h_3	q_1	q_2	q_3
gluons:	G_1		0		Q	$-Q$	0
	G_2				0	Q	$-Q$
monopoles:	M_1	R	$-R$	0	S	$-S$	0
	M_2	0	R	$-R$	0	S	$-S$
quarks:	ψ_1				$\frac{2}{3}Q$	$-\frac{1}{3}Q$	$-\frac{1}{3}Q$
	ψ_2		0		$-\frac{1}{3}Q$	$\frac{2}{3}Q$	$-\frac{1}{3}Q$
	ψ_3				$-\frac{1}{3}Q$	$-\frac{1}{3}Q$	$\frac{2}{3}Q$
condensed states	$M_1^3\overline{G}_1$	$3R$	$-3R$	0	$3S-Q$	$Q-3S$	0
	$M_2^3\overline{G}_2$	0	$3R$	$-3R$	0	$3S-Q$	$Q-3S$
liberated quarks	$\overline{M}_1^2\overline{M}_2\psi_1$	$-2R$	R	R	$\frac{2}{3}Q-2S$	$S-\frac{1}{3}Q$	$S-\frac{1}{3}Q$
	$\overline{M}_2 M_1\psi_2$	R	$-2R$	R	$S-\frac{1}{3}Q$	$\frac{2}{3}Q-2S$	$S-\frac{1}{3}Q$
	$M_1 M_2^2\psi_3$	R	R	$-2R$	$S-\frac{1}{3}Q$	$S-\frac{1}{3}Q$	$\frac{2}{3}Q-2S$

$Q=\frac{1}{2}g$, $R=4\pi/g$, $S=\theta g/4\pi$.

We stress that the phenomena described in this section will not occur in ordinary QCD where θ is observed to be very close to zero.

One can imagine a model in which the liberated quarks themselves, being bosons, develop vacuum expectation values. They could even be responsible in the first place for a system to condense into an oblique confinement mode. This observation will perhaps allow us to construct unusual models for interactions of elementary particles at extremely short distance. Is the presently popular Higgs particle a fermion-monopole bound state? We leave such questions for the future.

11. Conclusions

Non-abelian gauge theories can be cast in the form of abelian gauge theories with magnetic monopoles. We can distinguish "elementary" monopoles from bound monopole pairs. The elementary ones are those that arise as single singularities when the non-abelian part of the gauge is fixed by a non-propagating gauge condition. For SU(N) for instance there are only $N-1$ species, each carrying two opposite and consecutive U(1) charges according to (4.9). The others, such as those with two non-consecutive opposite charges, are bound states composed of elementary ones. Not only the abelian electromagnetic interactions but also the non-gauge interactions between these objects should be precisely determined by the microscopic theory. However, it may be that smoother gauge conditions are necessary such as (7.2) in which case both the ghost spectrum and the monopole spectrum become more complicated.

Whether or not this approach may lead to useful quantitative computation procedures remains to be seen. Abelian gauge theories with monopoles may be more transparent than non-abelian theories–they are still strongly interacting systems defying any conventional perturbative treatment. An advantage of our construction is that the dual transformation electric-magnetic now seems to be basically straightforward. However, since the fundamental gluons have spin one and the monopoles spin zero, and, moreover, since the monopoles come in fewer varieties than the gluons, the theory is as yet far from being self-dual.

One should also not underestimate a difficulty exposed by the instanton angle θ. If θ runs from zero to 2π the monopole changes into a monopole-gluon bound state. How then can we ever distinguish "pure" monopoles from monopole-gluon bound states? Such a distinction must be made on a rather arbitrary basis.

A byproduct of our considerations is the categorization of monopoles in the maximal abelian subgroup. This enabled us to enumerate the various possible phases in a gauge theory. Once θ is allowed to have non-trivial values the number of possible phases is much larger than realized before: they are defined by the choice of the linear subspace in the electric-magnetic charge lattice spanned by the condensed particles. Most of these "oblique confinement" modes are probably extremely difficult to realize in any conceivable model, but in principle they are possible. And all these modes will be separated by sharp phase transitions.

The author thanks A. Bais for explaining to him the nature of the phase transition in θ and P. van Baal for correcting an error in an original version of this paper. Further he had fruitful discussions with S. Gupta, C.P. Korthals Altes, S. Mandelstam, A.V. Manohar, I. Singer and N. Snyderman.

References

[1] B.S. DeWitt, Phys. Rev. 162 (1967) 1195, 1239;
L.D. Faddeev and V.N. Popov, Phys. Lett. 25B (1967) 29;
S. Mandelstam, Phys. Rev. 175 (1968) 1580, 1604
[2] G. 't Hooft, Nucl. Phys. B33 (1971) 173;
G. 't Hooft and M. Veltman, Nucl. Phys. B50 (1972) 318; DIAGRAMMAR, CERN Report 73/9 (1973)
[3] B.W. Lee, Phys. Rev. D5 (1972) 823;
B.W. Lee and J. Zinn-Justin, Phys. Rev. D5 (1972) 3121, 3137, 3155; D7 (1973) 1049;
E.S. Abers and B.W. Lee, Phys. Reports 9 (1973) 1
[4] H.B. Nielsen and P. Olesen, Nucl. Phys. B61 (1973) 45;
B. Zumino, *in* Renormalization and invariance in quantum field theory, ed. E.R. Caianiello (Plenum, New York, 1974) p. 367
[5] G. 't Hooft, Proc. E.P.S. Int. Conf. on High-energy physics, Palermo, Sicily, 23–28 June 1975
[6] G. 't Hooft, Nucl. Phys. B138 (1978) 1; B153 (1979) 141
[7] P.W. Higgs, Phys. Lett. 12 (1964) 132; Phys. Rev. Lett. 13 (1964) 508; Phys. Rev. 145 (1966) 1156;
F. Englert and R. Brout, Phys. Rev. Lett. 13 (1964) 321;
G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, Phys. Rev. Lett. 13 (1964) 585;
T.W.B. Kibble, Phys. Rev. 155 (1967) 627
[8] G. 't Hooft, Nucl. Phys. B35 (1971) 167
[9] A.M. Polyakov, JETP Lett. 20 (1974) 194;
G. 't Hooft, Nucl. Phys. B79 (1974) 276;
F. Englert and P. Windey, Phys. Rev. D14 (1976) 2728;
P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. B125 (1977) 1
[10] P.A.M. Dirac, Proc. Roy. Soc. A133 (1931) 60; Phys. Rev. 74 (1948) 817
[11] E. Witten, Phys. Lett. 86B (1979) 283
[12] R. Jackiw and C. Rebbi, Phys. Rev. Lett. 36 (1976) 1116;
P. Hasenfratz and G. 't Hooft, Phys. Rev. Lett. 36 (1976) 1119;
A. Goldhaber, Phys. Rev. Lett. 36 (1976) 1122
[13] G. 't Hooft, Acta Phys. Austr., Suppl. XXII (1980) 531
[14] G. 't Hooft, Lectures at Symp. on Perspectives in modern field theories, Stockholm, September 1980, Phys. Scripta, to be published