

## TOPOLOGICAL SYMMETRY BREAKDOWN AND QUARK CONFINEMENT \*

Stuart SAMUEL

*Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720, USA*

Received 12 February 1979

The manner by which topological symmetry breakdown leads to confinement in a non-Abelian gauge theory is discussed.

### 1. Introduction

In a recent paper [1], 't Hooft has stressed the importance of the center of the group in non-Abelian gauge theories. This has exposed new concepts and created new possibilities for quark confinement. The center of  $SU(N)$  is  $Z(N)$ , isomorphic to the set of integers,  $0, 1, 2, \dots, N-1$ , under the operation of addition modulo  $N$ .  $Z(N)$ , with its finite number of elements and unusual modulo addition property, is a feature distinguishing non-Abelian theories from Abelian ones. It may be the crucial factor explaining why non-Abelian theories confine though Abelian theories do not. Instantons also differentiate  $SU(N)$  gauge theories. However, dilute instanton gases do not confine and dense instanton gases are hard to handle. The runaway scale and density problem has made instanton calculations virtually impossible to do. There are now conflicting views [2,3] on their relevance to physical processes, problems and confinement. 't Hooft  $Z(N)$  type excitations, on the other hand, have only been discussed in a formal manner. The calculations remain to be done and it is unknown whether they will encounter similar difficulties. Thus, this is an important area of research. This is what we will be discussing in this paper. In particular, we will show how some calculations, such as Wilson loop integrals, can be done in a manner similar to instanton ones. We will also discuss many physically interesting ideas although no computations will be done to support them.

Unlike 't Hooft, who used a Hamiltonian approach, we shall use a Euclidean formulation. We find that the properties of  $Z(N)$  excitations are particularly simple from this point of view, especially when considering Wilson loops, where the effect of the excitations is expressed in terms of linking numbers.

\* This work has been supported by the High Energy Physics Division of the United States Department of Energy under contract no. W-7405-ENG-48.

Although 't Hooft had suggested attacking the problem by going to a lattice and many people have begun considering  $Z(N)$  lattice gauge theories [4], we shall work in the continuum.

't Hooft discussed, in detail, but in a formal way, the nature of confinement in  $2 + 1$  dimensions. Here is a review of how it works. One starts with an  $SU(N)$  gauge theory. Using a symmetry breaking Higgs potential,  $SU(N)$  is broken down to  $Z(N)$ . Topological solitons can then occur. These are not so different from Nielsen-Olesen vortices [5] except that the non-zero gauge potentials are proportional to  $\lambda^8$  (for example) in the  $SU(3)$  case and more Higgs fields are involved. Away from a Nielsen-Olesen vortex, field configurations look like a gauge transformation,  $U(x)$ , with  $U(x) = \exp(i\phi)$ , so that  $A_\mu = -(i/e)[\partial_\mu U(x)]U^{-1}(x)$  and  $\Phi(x) = U(x)F$  (see sect. 3 for notation). Likewise, 't Hooft vortices are approximately singular gauge transformations,  $U_{x_0}(x)$  ( $x_0$  is the location of the soliton and  $x$  is the point where the gauge transformation is applied). When  $U_{x_0}(x)$  is written as  $\exp[i\sum_{i=1}^8 \alpha_i(x) \frac{1}{2}\lambda_i]$ , that is, the fundamental matrix representation is used,  $U_{x_0}(x)$  has the property that when  $x$  encircles  $x_0$  and returns to  $x$ ,  $U_{x_0}(x)$  is an element of the center of the group. Such a transformation is multivalued (and globally ill-defined), however, when written in the adjoint representation (for example), it becomes single-valued. In 't Hooft's model, gauge and Higgs fields are in the adjoint representation so that  $U_{x_0}(x)$  is well-defined. If  $\langle \Phi_i(x) \rangle = F_i$  are a set of vacuum expectation values which minimize the Higgs potential, then the classical configuration for an 't Hooft  $Z(N)$  soliton at  $x_0$  is

$$A_\mu = -\frac{i}{g} [\partial_\mu U_{x_0}(x)] U_{x_0}^{-1}(x),$$

and  $\Phi_i(x) = U_{x_0}(x)F_i$  far from the vortex. Of course, nearby the field configuration is non-trivial just as in the Nielsen-Olesen case.

In  $2 + 1$  dimensions the above solitons are particles with extended structure and non-zero form factors. They are stable for topological reasons and carry a topological conserved charge. This charge is governed by the group,  $Z(N)$ , which means that charge is conserved modulo  $N$  so that, in principle,  $N$  charges may annihilate. This  $Z(N)$  group is completely different from the one associated with the center of  $SU(N)$  and should not be confused with it. There are now two  $Z(N)$ 's. The one associated with the soliton charge will be called topological  $Z(N)$ . 't Hooft argues that it may be possible that topological  $Z(N)$  is spontaneously broken, a phenomenon we call topological symmetry breakdown. This is an interesting phase in which quark confinement occurs. The argument is as follows: if topological  $Z(N)$  is broken, then there are  $N$  different kinds of vacuums characterized by their topological  $Z(N)$  numbers (compare this to spontaneous symmetry breakdown of a  $U(1)$  symmetry by a Higgs potential, where, instead, there is a continuum (a circle) of vacuums defined by the direction in which the Higgs field points). In general, at any instant in time, the physical vacuum will look like a collection of domains each characterized by its  $Z(N)$  value. Separating these domains will be Bloch walls. They carry an energy per

unit length and may be associated with a new quanta in the theory: closed strings. When particles in the fundamental representation are introduced, 't Hooft argues that they will be confined. A quark and an antiquark will have a Bloch-wall-like string between them. This will provide a linear confining potential. For  $SU(3)$  three strings may join so that baryons can consist of three confined quarks. This is 't Hooft's  $2 + 1$  dimensional quark confinement scheme. He derived it using simple, physical, intuitive arguments.

Several questions are generated. First is: how does one extend these notions to  $3 + 1$  dimensions? Following the same line of reasoning there will be volumes of  $Z(N)$  vacua (instead of areas) and closed surfaces separating them (instead of closed strings). Clearly closed surfaces are unable to interpolate between quarks and create a linear potential as in one lower dimension. For this reason 't Hooft conjectured that confinement in  $3 + 1$  dimensions is different. Instead of proposing a confinement scheme, he settled for an operator algebra which allowed the determination of the different phases of the theory. One of the important results of our paper will be the extension of the  $2 + 1$  dimensional scheme to  $3 + 1$  dimensions. The reason we are able to do this is that in a Euclidean formulation of the  $2 + 1$  dimensional model we find a slightly different picture of the confinement, which has a straightforward generalization to one higher dimension.

A second question is: how does one do calculations? 't Hooft has used formal powerful arguments, but it remains an open problem as to how to do computations. Of particular interest is the coefficient in front of the linear potential and its companion, the slope parameter. More generally, how does one calculate in a theory with topological symmetry breakdown? We are able to supply some of the answers. We treat the Nielsen-Olesen case in sect. 3 and the 't Hooft model in sect. 4. To do these calculations requires a new calculation method. We develop it in sect. 2. The problem is equivalent to treating a gas of closed loops. Similar problems arise in lattice field theories. Examples are the three-dimensional  $O(2)$  classical Heisenberg model and the four-dimensional Abelian lattice gauge theory discussed by Banks, Myerson and Kogut [6]. These authors must deal with a gas of monopole loops. Our gas consists of Nielsen-Olesen or 't Hooft vortex loops. We treat such a system by developing a field theory to describe it. Very simple arguments then tell us about the quark-antiquark and three-quark potentials. Although our main interest is in topological symmetry breakdown and Wilson loop calculations, sect. 2 discusses several field theory phenomena in this new picture. These concepts are enumerated and briefly discussed.

A third question is: what are the essential ingredients in 't Hooft's  $2 + 1$  dimensional confinement scheme? Certainly topological symmetry breakdown is one of them but are there others? We find the answer is yes. Topological symmetry breakdown leads only to a logarithmic quark-antiquark potential unless another ingredient is also present. It is the monopole. The 't Hooft model has monopoles in it. They play an instrumental role in the quark confinement. Before topological symmetry breakdown, the monopoles are bound together in monopole-antimonopole

pairs. These dipoles have little effect. Topological symmetry breakdown liberates these monopoles. They, in turn, confine charges in a manner not so different from Polyakov [7] and Mandelstam [8]. In sect. 7, we relate these ideas to Mandelstam's quark confinement scheme. The  $Z(N)$  confinement in  $3 + 1$  dimensions is practically the same as Mandelstam's. We consider this to be an important result: two seemingly different confinement schemes are, in fact, the same.

The remaining open problems and questions (and there are several) are presented at the end.

## 2. Closed-loop gas as a field theory

This section will relate a gas of closed continuous loops to a relativistic field theory\*. Connections between statistical mechanics and field theory often prove useful [10]. We find this to be the case here and will use it to extract results in a physical and almost intuitive manner.

We shall proceed in steps. Note, first, however, that a continuous curve, when broken into  $N$  segments, resembles a polymer with vertices acting as atoms and line segments acting as bonds (fig. 1). Sometimes this analogy is useful. Consider an open macromolecule (or polymer) which goes from  $x_0$  to  $x_N$  via  $x_1, x_2, \dots, x_{N-1}$  (fig. 1b). To enforce the condition that the curve be continuous, we demand that the  $i$ th atom be near its two neighbors. This can be done by requiring two neighboring atoms to be a distance  $\epsilon$  from each other, that is, the bonds have a fixed length,  $\epsilon$ . The total length is  $N\epsilon$ . Allowing curves of different lengths means summing over  $N$  in the partition function. To make the model more physical, assume the atoms have a chemical potential,  $\mu$ , and interact with a "strength"  $g$  to an external potential,  $V(x)$ . The grand partition function for a macromolecule with ends fixed at  $x_0$  and  $x_f \equiv x_N$  is

$$\begin{aligned}
 Z(x_0, x_f) &= \sum_{N=1}^{\infty} \exp(-\beta\mu N) Z_N(x_0, x_f), \\
 Z_N(x_0, x_f) &= \left[ \prod_{i=1}^{N-1} \int d^3x_i \right] \prod_{i=1}^N \left[ \frac{\delta(|x_i - x_{i-1}| - \epsilon)}{4\pi\epsilon^2} \right] \\
 &\quad \times \exp\left[-\beta \sum_{i=1}^N gV(x_i)\right].
 \end{aligned} \tag{2.1}$$

$Z_N$ , the partition function for a macromolecule with  $N + 1$  atoms and  $N$  bonds, is

\* Stone and Thomas have used such an analogy in ref. [9]. We prefer to rederive their result in a lattice-independent way.

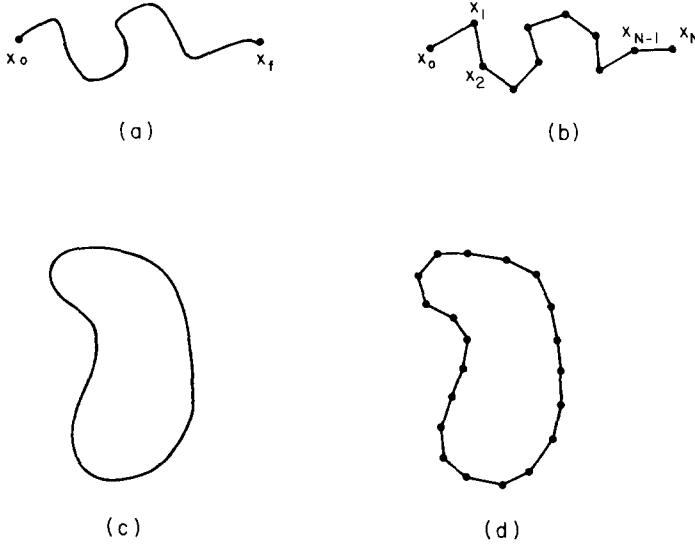


Fig. 1. The macromolecule approximation. (a) A continuous curve between  $x_0$  and  $x_f$ , (b) the macromolecule approximation of (a), (c) a closed curve, and (d) its macromolecule approximation.

a summation (integration) over all positions of intermediate atoms weighted by Boltzmann factors,  $\exp[-\beta g V(x_i)]$ , with the constraint (the delta functions) that bonds have length,  $\epsilon$ . The factor,  $1/4\pi\epsilon^2$ , normalizes this so that given the position of the  $i$ th atom, the integration over the location of the  $(i+1)$ th atom is unity. Modification of this factor can always be absorbed in  $\mu$ . As long as  $x_f$  and  $x_0$  are far apart and  $\epsilon$  is small, the sum over  $N$  effectively begins with the enormous value,  $N_0 = |x_f - x_0|/\epsilon$ . It does no harm to start the sum at  $N = 1$ .

Eventually, we will take  $\epsilon$  to zero so as to recover continuous curves from segmented  $N$ -step ones. In this limit,  $Z_N$  resembles a Feynman path integral. Path integrals have been widely used to account for the statistical properties of macromolecules [11]. We shall review this. As is common in statistical mechanics, one must coarse-grain: write  $N$  as  $nm$  with both  $n$  and  $m$  large, that is, break up the macromolecule into  $m$  units of  $n$  atoms. Consider the situation where  $\epsilon$  is small,  $n$  is large, but  $\sqrt{n}\epsilon$  is also small. Then

$$\prod_{i=vn}^{(v+1)n-1} \frac{1}{4\pi\epsilon^2} \delta(|x_i - x_{i-1}| - \epsilon) \rightarrow \left(\frac{3}{2\pi n\epsilon^2}\right)^{3/2} \times \exp\left[-\frac{3}{2n\epsilon^2} (x_{(v+1)n} - x_{vn-1})^2\right]. \quad (2.2)$$

In eq. (2.2)  $i = \nu n$  to  $(\nu + 1)n - 1$  are the atoms in the  $\nu$ th unit. Eq. (2.2) is true because the left-hand side represents a random walk between  $x_{\nu n-1}$  and  $x_{(\nu+1)n}$  which by the Central Limit Theorem approaches a Gaussian. If  $\sqrt{n}\epsilon$  is small compared to the distance over which  $V(x)$  varies appreciably, then  $V(x)$  may be treated as a constant in each unit. We obtain

$$Z = \sum_{N=1}^{\infty} \left[ \prod_{\nu=1}^{m-1} d^3 x_{\nu} \right] \left[ \left( \frac{3}{2\pi n \epsilon^2} \right)^{3m/2} \exp \left\{ -\frac{3}{2} \sum_{\nu=1}^m \frac{(x_{\nu} - x_{\nu-1})^2}{n \epsilon^2} \right\} \right] \\ \times \exp \left[ -\beta \sum_{\nu=1}^m n V(x_{\nu}) - \beta \sum_{\nu=1}^m n \mu \right]. \quad (2.3)$$

The  $x_i$ 's have been relabelled so that  $x_{\nu}$  is the average value of  $x$  in the  $\nu$ th cell. As  $\epsilon$  goes to zero, eq. (2.3) approaches a Feynman integral. Using the variables

$$s \equiv \frac{1}{6} n \epsilon^2 \nu, \quad \tau \equiv \frac{1}{6} \epsilon^2 N, \quad (2.4)$$

and the "bare" mass and coupling defined by

$$m_0^2 = \frac{6\beta\mu}{\epsilon^2}, \quad g_0 = \frac{6g}{\epsilon^2}, \quad (2.5)$$

eq. (2.3) becomes

$$Z = \frac{6}{\epsilon^2} \int_0^{\infty} d\tau \int_{\substack{x(0)=x_0 \\ x(\tau)=x_f}} \mathcal{D}x \exp \left[ -\int_0^{\tau} \left( \frac{1}{4} \dot{x}^2(s) + m_0^2 + \beta g_0 V(x(s)) \right) ds \right]. \quad (2.6)$$

The sum over  $N$  has become an integral over  $\tau$ .

*Remarks:* (a) A bond-vector field interaction of the form

$$\sum_i q(x_i - x_{i-1})_{\mu} H^{\mu}(\bar{x}_i)$$

can also be considered ( $\bar{x}_i \equiv \frac{1}{2}(x_i + x_{i-1})$  is the average value of  $x$  along the bond). The effect is to add a term  $-\beta q \int_0^{\tau} \dot{x}_{\mu}(s) H^{\mu}(x(s)) ds$  to the action in eq. (2.6).

(b) Eq. (2.6) resembles the Green functions which arise in a particle dynamics representation of a field theory [12].

(c) The factor  $6/\epsilon^2$  can be removed by normalizing  $Z$  appropriately (a sort of wave-function renormalization). In any case, physical quantities (averages) do not depend on the overall normalization of  $Z$ .

Summarizing, *the Green functions,  $G(x_f, x_0)$ , which appear in a particle dynamics representation of a field theory, correspond to the grand partition functions of polymers with ends at  $x_f$  and  $x_0$ . The sum is over all continuous paths of arbitrary shape and length. The mass squared is proportional to the chemical potential which must*

be appropriately scaled to obtain a continuum limit. Other inputs are (a) that  $n \rightarrow$  large so that products of delta functions approach Gaussians, (b) that  $\sqrt{n} \epsilon \rightarrow 0$  so that  $\sum_{x_i \in \text{unit}} V(x_i) \rightarrow nV(\bar{x})$ , and (c) that  $\epsilon \rightarrow 0$  so that segmented curves approach smooth ones.

Representations such as eq. (2.6) are well-known to chemical physicists [11]. Eq. (2.6) is not a new result. We have rederived it as a warm-up for the next step: polymers in bulk. It is not hard to believe that a gas of polymers might be describable as a Euclidean field theory, and we shall show this. We have known about this correspondence for some time and have thought about using it to obtain field theory results as in ref. [10]. Until the present application, this analogy did not seem fruitful because a gas of closed loops is a complicated statistical system of which little is known. Fortunately, we shall not use the statistical mechanics side of this analogy in an essential way.

Now consider closed polymers, obtained by setting  $x_f = x_0 \equiv x$  in eq. (2.6). Allowing loops to be located anywhere necessitates integrating over  $x$ . There is, however, an overcounting problem. For a closed polymer of  $N$  atoms, it is impossible to differentiate which atom was the starting point, that is, the  $N$  different starting points cannot be distinguished. For each configuration that begins at  $x_0$ , traces an  $N$ -fold segmented path and returns to  $x_0$ , there is one which begins at  $x_i$ , traces out the self-same path, and returns to  $x_i$ . Thus configurations are overcounted by a factor of  $N$ . The partition function for a closed polymer of arbitrary length and location is

$$Z = \int d^3x \sum_{N=1}^{\infty} \frac{1}{N} \exp[-\beta\mu N] Z_N(x, x), \quad (2.7)$$

where  $Z_N$  is given in eq. (2.1). Proceeding as before, we obtain

$$Z = \int_0^{\infty} \frac{d\tau}{\tau} \int d^3x \int_{\substack{x(0)=x \\ x(\tau)=x}} \mathcal{D}x \exp[-\int_0^{\tau} (\frac{1}{4}\dot{x}^2 + m_0^2 + \beta g_0 V(x)) ds]. \quad (2.8)$$

Finally, the grand partition function for the gas of loops is

$$\begin{aligned} \mathcal{Z} &= \sum_{M=0}^{\infty} \frac{1}{M!} Z^M = \exp Z \\ &= \mathcal{N} \exp\{-\text{tr} \ln[p^2 + m_0^2 + \beta g_0 V]\} \\ &= \mathcal{N} \iint \mathcal{D}\phi \exp\{-\int [\partial_{\mu}\phi \partial^{\mu}\phi(x) + m_0^2 \phi^2(x) + \beta g_0 V(x) \phi^2(x)] d^3x\}. \quad (2.9) \end{aligned}$$

$\mathcal{N}$  is an overall (infinite) normalization constant. We have used

$$\int_0^{\infty} \frac{d\tau}{\tau} [\exp(-a\tau) - \exp(-b\tau)] = \int_a^b dz \int_0^{\infty} \exp(-z\tau) d\tau = -\ln \frac{a}{b}.$$

The ill-defined innocuous extra piece,

$$\exp\left[-\int d^3x \int_0^\infty \frac{d\tau}{\tau} \exp(-b\tau)\right],$$

is  $\mathcal{N}$ . The operator,  $i\partial_\mu$ , is  $p_\mu$  and  $\phi$  is a scalar field.

*Remarks:* (a) As usual, the infinity in  $\mathcal{N}$  is harmless since it divides out when calculating expectation values.

(b) For oriented curves (that is, curves with a direction for which curves of different direction are distinguishable) a functional integral over a complex (charged) field is obtained. The orientation direction is identified with the flow of charge. In general, a gas of  $T$  different types of macromolecules leads to a  $T$ -component field.

(c) With oriented curves and a bond interaction, the action,

$$A = \int [i(\partial_\mu - ieA_\mu)\phi]^2 + m_0^2\phi^*\phi] d^3x,$$

can be obtained.

(d) Interactions between atoms (and/or bonds) can be introduced using auxiliary fields. Suppose the interactions are of the form

$$\sum_{\substack{\text{all pairs} \\ \text{of atoms}}} g^2 V(x_i, x_j).$$

Define

$$\int G(x, y) V(y, z) d^3y = \delta^3(x - y).$$

Then

$$\begin{aligned} \mathcal{Z} &= \frac{1}{\iint \mathcal{D}\chi \exp\{\frac{1}{2}\beta \int \chi(x) G(x, y) \chi(y) d^3x d^3y\}} \\ &\times \iint \mathcal{D}\chi \iint \mathcal{D}\phi \exp\{-\int [\partial_\mu \phi \partial_\mu \phi + m_0^2 + \beta g_0 \chi(x)] d^3x\} \\ &\times \exp\{\frac{1}{2}\beta \int \chi(x) G(x, y) \chi(y) d^3x d^3y\}. \end{aligned} \tag{2.10}$$

This is verified by first doing the  $\phi$  integration to yield a gas of closed loops and then by doing the  $\chi$  integration.

(e) In particular, a  $\phi^4$  theory corresponds to a repulsive interatomic delta-function potential.

(f) Three-dimensional scalar QED corresponds to a bond-bond interaction of the form

$$\frac{e^2}{4\pi\beta} \sum_{\substack{\text{pairs of} \\ \text{bonds,} \\ \mathbf{b} \text{ and } \mathbf{b}'}} \frac{\mathbf{b} \cdot \mathbf{b}'}{r_{\mathbf{b}\mathbf{b}'}}. \tag{2.11}$$



In eq. (2.11)  $\mathbf{b} = \mathbf{x}_{i+1} - \mathbf{x}_i$  is the bond vector between the  $i$ th and  $(i + 1)$ th atoms,  $r_{bb'}$  is the distance between the two bonds, and the sum is over all pairs of bonds. The interaction is attractive for antiparallel bonds and repulsive for parallel ones.

(g) The Lagrangian,

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m_0^2 \phi^2 - \beta g_0 \chi \phi^2 - \frac{1}{2}\beta(\partial_\mu \chi)^2 - \frac{1}{2}\beta m_\chi^2 \chi^2,$$

corresponds to attractive interatomic Yukawa potentials. When  $m_\chi = 0$ , Coulomb interactions are obtained.

(h) The method works in any dimension, of course.

(i) Our method can be applied to obtain local field theories for solitons [9,13]. As ref. [13] points out, perturbative expansion about vacuum expectation values misses soliton solutions. These extra solitons carry topological charges which are conserved. They generate closed loops in the appropriate dimension, *via* a macromolecule analogy yield partition functions similar to  $Z_N(x, x)$  of eq. (2.1) and  $Z$  of eq. (2.7), and therefore result in a field theory. Of course, the solitons interact with the original fields. This can be taken into consideration using auxiliary fields and Lagrange multipliers (as done in ref. [13]). The macromolecule technique would replace the left-hand side of eq. (3.3) of ref. [13] by

$$\sum_{M=0}^{\infty} \frac{1}{M!} Z^M, \quad (2.12)$$

with

$$Z = \int_0^\infty \frac{d\tau}{\tau} \int d\mathbf{x} \iint_{\substack{\mathbf{x}(0)=\mathbf{x} \\ \mathbf{x}(\tau)=\mathbf{x}}} \mathcal{D}\mathbf{x} \exp\left[-\int_0^\tau \left(\frac{1}{4}\dot{\mathbf{x}}^2(s) + m_0^2 + iQ_\mu(x(s))\dot{x}^\mu(s)\right) ds\right]. \quad (2.13)$$

This would provide an alternative derivation to the one of appendix A of ref. [13]. Although the normalization of delta functions in eq. (2.1) is unknown and could generate an unknown mass parameter,  $m_0^2$ , our intuitive feeling is that  $m_0^2 = 0$ , although we cannot prove it. If it were non-zero, the measure factors in going to a sum over trajectories, would determine it in terms of the parameters in the original Lagrangian. One might think that  $m_0^2$  could be the missing factor cancelling self-energy infinities found in ref. [13] (sect. 4), but we believe this is not the case. The infinities occur because near the soliton the Higgs field must go to zero. Expanding about non-zero Higgs field vacuum expectation values cannot deal with such a constraint. Regardless, the mass,  $m_0$ , is geometrical in nature and has nothing to do with the soliton's physical mass.

(j) The chemical potential per atom and  $m_0^2$  are proportional. For  $m_0^2 < 0$ , the chemical potential is negative and loops will populate the vacuum indefinitely. When a symmetry breaking potential is used, repulsive delta-function potentials between atoms stabilize the proliferation of loops. Therefore, spontaneous symmetry breakdown looks like a dense gas of loops from the symmetric vacuum point of view (the "spaghetti vacuum").

(k) Renormalization infinities occur because the macromolecule potentials corresponding to interacting relativistic field theories are singular. For example, a  $\phi^4$  theory corresponds to repulsive delta-function potentials between atoms. This extremely short-ranged singular potential consequently ruins the coarse-graining procedure used in going from eq. (2.1) to (2.6). It will no longer be true that these “random walks” approach Gaussian distributions. A  $\phi^4$  theory is not so different from the self-avoiding random-walk problem [11,14]. The non-Gaussian nature of this process is well-known [14]. This will lead to wave-function and mass renormalization as perturbation theory tries to approximate non-Gaussian processes by Gaussian ones. The singular nature of potentials can also cause other problems. For attractive potentials, the interatomic forces might be too strong (“non-renormalizability”) and cause macromolecules to collapse into “balls of wire”. Perturbation theory is insensitive to the sign of  $g$  and therefore such effects manifest themselves for repulsive potentials, also. The higher the dimension of space-time, the more singular the forces. This is why fewer renormalizable theories occur in higher dimensions.

### 3. Wilson loops in the presence of topological vortices

This section will calculate the Wilson loop in the presence of a gas of Nielsen-Olesen vortices [5]. Sect. 4 will treat 't Hooft  $Z(3)$  vortices. These calculations are similar to instanton calculations, where multiple instanton configurations generate a gas. Various computational devices have put instanton calculations on a solid foundation [7]. Statistical mechanics and physical intuition determine their properties quite easily. This is how Callan, Dashen and Gross [2] are able to determine the magnetic properties of a dense BPST [15] instanton gas. Contrast this with a vortex gas. The vortices may vary in number, position and the way they are imbedded. They may twist in the most unruly manner. These complications lead one to think that such a system is too difficult to deal with: however, the methods of sect. 2 make the problem tractable. We are able to do Wilson loop calculations. In fact, gases of vortices are as easy to handle as gases of instantons. We will show how calculations involving topological spontaneous symmetry breakdown are done.

Take the Euclidean space version of the 2 + 1 dimensional Nielsen-Olesen model [5]. This is scalar QED with a Higgs potential. The Lagrangian is

$$\mathcal{L} = \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |(\partial_\mu - ieA_\mu)\Phi|^2 + \lambda(\Phi^*\Phi - F^2)^2, \quad (3.1)$$

and has vortex-like solutions along the third axis of the form

$$\Phi(x) = f(\rho) \exp(i\phi), \quad A_\phi(x) = a(\rho), \quad A_z(x) = A_\rho(x) = 0. \quad (3.2)$$

Cylindrical coordinates,  $\rho, z, \phi$ , have been used. Graphs of  $f(\rho) = |\Phi(\rho)|$  and  $H_z(\rho)$  are given in fig. 1 of ref. [5].

The important properties of the solution are:

- (i)  $(\Phi_1, \Phi_2)$  points radially outward from the vortex;
- (ii)  $|\Phi(\rho)|$  vanishes at the vortex (at  $\rho = 0$ ) and goes to  $F$  far from it ( $\rho = \infty$ );
- (iii)  $a(\rho) \rightarrow 1/e\rho$  for from the vortex. This means that the vortex contains a tube of magnetic flux. The total flux is

$$\int dx dy H_z(x, y) = \int_0^{2\pi} A_\phi(\rho)\rho d\phi = 2\pi/e .$$

Property (i) is topological in nature and makes the soliton topologically stable.

Property (ii) has important implications (short-ranged ones and long-ranged ones). First, near the vortex where  $|\Phi(\rho)| = 0$ , the photon is effectively massless, whereas far away it has mass because the Higgs field has a vacuum expectation value. This is one way of understanding why flux is channelled into tubes. There is a tubular “mass-confinement” bag. Also  $|\Phi(\rho = 0)| = 0$  indicates symmetry restoration in the vortex, a point which we shall discuss in detail later. Finally, we should mention that the Higgs field prevents the vortex from collapsing to zero size and gives it a finite mass. Secondly, the Higgs field at infinity must take on vacuum expectation values and be covariantly constant. The latter means that  $A_\phi(\rho)$  is determined in terms of the phase,  $\chi$ , of the Higgs field.  $A_\mu$  is pure gauge and  $\chi(x) - \chi(y) = e \int_x^y A_\mu dx^\mu$ . Whenever  $y$  loops around a circle and returns to  $x$  the phase,  $\chi$ , must be an integral multiple of  $2\pi$ . This causes the flux to be quantized. We shall return to this point in sect. 8.

Property (iii) contains the physics: the vortices are quantized tubes of magnetic flux. This is the key physical characteristic.

The vortex soliton has a topological number, which can be seen in two ways: using gauge potentials or using Higgs fields\*. Of course, the two are interrelated. In terms of  $A_\mu$ , consider  $\exp[ie \int_x^y A \cdot dx]$ . Such a pure phase takes values on the unit circle in the complex plane. Fix  $x$  and move  $y$  around the vortex as in fig. 2. In regions where  $A_\mu$  is pure gauge and the Higgs field is covariantly constant,  $\exp[ie \int_x^y A \cdot dx]$  must return to 1 when  $y$  returns to  $x$ . This forms a map from a circle in Euclidean space to the unit complex circle. These maps are characterized by winding numbers,  $\pi_1(S^1) = \mathbb{Z}$ , the set of integers. Winding numbers count the number of times the image curve loops around the unit circle. Regions of space with  $\pi_1 = 1$  or  $\pi_1 = -1$  contain a vortex or an antivortex.

In terms of the Higgs field the topology is as follows: far away  $|\Phi|$  must be  $F$ , that is,  $\Phi(x)$  must take on values which minimize the potential,  $\lambda(\Phi^* \Phi - F^2)^2$ ; they form a circle of minima. Again consider moving  $y$  around a closed curved (fig. 2). Then  $\Phi(y)$  forms a map from a circle in Euclidean space to a circle (of minima). Again these maps are characterized by  $\pi_1(S^1) = \mathbb{Z}$ .

\* This review of vortex topological properties is similar to that of Goddard and Olive who elucidate the topological properties of monopoles in ref. [16].

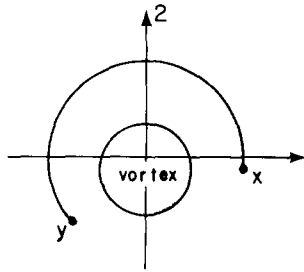


Fig. 2. A vortex. This is a cross section of the vortex in the 1–2 plane. The 3rd axis is out of the paper.

Because winding number is invariant under continuous deformations (homotopies), winding number can neither be created nor destroyed unless an (unallowed?) discontinuity occurs in field configuration. This provides the topological stability of the vortices.

In three-dimensional space-time, orient the vortex lines. The orientation indicates the direction of the flux and the flow of topological charge. In this way antivortices

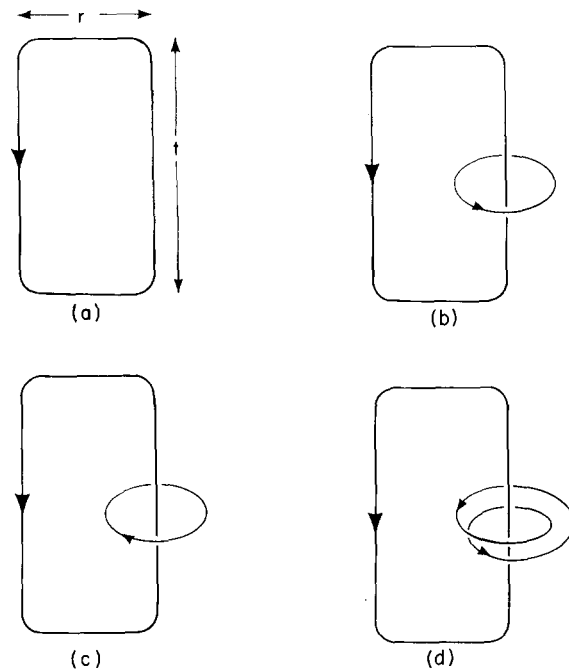


Fig. 3. Linking numbers. (a) The Wilson loop. Its width and length are  $r$  and  $t$ . (b), (c) and (d) show a vortex linking with this Wilson loop with linking numbers  $+1$ ,  $-1$  and  $+2$ , respectively.

act like particles travelling backward in time. Of course, vortices need not be straight lines. In general, they twist and curve in an arbitrary manner. Since they may never end, they form closed loops. If they have a positive mass, these loops will be small, will act like neutral objects, and will have few physical consequences. In contrast, when their mass is negative, they fill up the vacuum and their effects can be dramatic

Having reviewed the properties of Nielsen-Olesen vortices, let us put a Wilson loop (with an arbitrary charge,  $q$ ) in the system (see fig. 3a):

$$\langle \exp[iq \oint A \cdot dy] \rangle. \quad (3.3)$$

□

Consider a single vortex. What is its effect? Using Stoke's theorem (or simply physical intuition), one sees that the Wilson loop measures the linking number [17] of the two curves. If the vortex links  $n$  times the contribution is  $\exp[i(2\pi/e)qn]$ . Examples with  $n = +1, -1$ , and  $+2$  are shown in figs. 3b, c and d. For several vortices the contribution is  $\exp[i(2\pi q/e)(n_+ - n_-)]$ , where  $n_+$  and  $n_-$  are the number of positive and negative linkages. Idealize the situation to the case where vortices are thin (trace out curves rather than tubes) and are not mutually interacting. This can be rectified if the form of the interactions is known. For simplicity set the mass equal to zero. In the end, we will restore a non-zero mass. The linking number of two curves can be expressed as an integral [17]:

$$n = -\frac{1}{4\pi} \int_{C_1} d\mathbf{x} \cdot \int_C \frac{(\mathbf{x} - \mathbf{y}) \times d\mathbf{y}}{|\mathbf{x} - \mathbf{y}|^3}. \quad (3.4)$$

Here,  $n$  is the number of times  $C_1$  and  $C$  link. For several curves the total linking number with  $C$  is

$$n = \sum_i -\frac{1}{4\pi} \int_{C_i} d\mathbf{x}_i \cdot \int_C \frac{(\mathbf{x}_i - \mathbf{y}) \times d\mathbf{y}}{|\mathbf{x}_i - \mathbf{y}|^3} = \sum_i \int_{C_i} d\mathbf{x}_i \cdot \mathbf{B}(\mathbf{x}_i), \quad (3.5)$$

where

$$\mathbf{B}(\mathbf{x}) = -\frac{1}{4\pi} \int_C \frac{(\mathbf{x} - \mathbf{y}) \times d\mathbf{y}}{|\mathbf{x} - \mathbf{y}|^3}. \quad (3.6)$$

Note that  $\mathbf{B}(\mathbf{x})$  has the same form as the magnetic field produced by a current flowing in  $C$ . Take  $C$  to be the Wilson loop and think of the  $C_i$  as vortex loops. Using the method of sect. 2, the calculation of (3.3) in a gas of idealized Nielsen-Olesen vortices is

$$\langle \exp[iq \oint A \cdot dy] \rangle = \sum_{M=0}^{\infty} \frac{1}{M!} \prod_{i=1}^M \left[ \int_0^{\infty} \frac{d\tau_i}{\tau_i} \int d\mathbf{x}_i \int_{\substack{x^i(0)=x_i \\ x^i(\tau_i)=x_i}} \mathcal{D}x^i \right]$$

$$\begin{aligned} & \times \exp \left\{ - \int_0^{\tau_i} \left( \frac{1}{4} (\dot{x}^i)^2 + \frac{2\pi i q}{e} B^\mu(x^i) \dot{x}_\mu^i \right) ds^i \right\} \Big/ [\text{same term with } B_\mu = 0] \\ & = \frac{\iint \mathcal{D}\psi \exp \{ - \int |(\partial_\mu - (i2\pi q/e) B_\mu) \psi|^2 \}}{[\text{same term with } B_\mu = 0]} . \end{aligned} \quad (3.7)$$

If the vortices interact and have a mass, eq. (3.7) would be replaced by

$$\frac{\iint \mathcal{D}\psi \exp \{ - \int (|(\partial_\mu - (i2\pi q/e) B_\mu) \psi|^2 + m^2 \psi^* \psi + \mathcal{V}(\psi^* \psi)) \}}{[\text{same term with } B_\mu = 0]} . \quad (3.8)$$

$\int \mathcal{V}(\psi^* \psi)$  might be, for example,

$$\int d^3x \int d^3y \psi^*(x) \psi(x) \frac{\exp[-\mu|x-y|]}{4\pi|x-y|} \psi^*(y) \psi(y).$$

This would correspond to a Yukawa interaction between points on the vortex. In any case  $\mathcal{V}$  must be a functional of  $\psi^*(x) \psi(x)$  only. Given the interactions and mass of vortices, the exact form of eq. (3.8) can be determined. In principle,  $m^2$  and  $\mathcal{V}(\psi^* \psi)$  may be approximately deduced using semiclassical methods or perhaps by examining the Lagrangians of ref. [13] in more detail.

Of course  $\psi$  is the vortex field. In fact, inserting  $\psi^*(x) \psi(y)$  in the integrand of eq. (3.8) (with  $B_\mu = 0$  so that the Wilson loop is absent) and returning to a macro-molecule description, one obtains a gas of closed vortices in the presence of one open one which starts at  $x$  and ends at  $y$ . Hence,  $\psi^*(x)$  produces the vortex endpoint at  $x$  and  $\psi(y)$  destroys it at  $y$ .

Regardless of the detailed nature of  $\mathcal{V}(\psi^* \psi)$ , we can evaluate eq. (3.8) semi-classically for various cases. (i) When  $m^2 > 0$ ,  $\psi = 0$  is expected to be the vacuum. This is also the solution to the equations of motion for a non-zero  $B_\mu$  and the Wilson loop to this approximation is 1. Vortices do not contribute to the Wilson loop as expected. This is because for  $m^2 > 0$  the vortices are small loops and rarely link with the Wilson loop. (ii) When  $m^2 < 0$ , there is topological symmetry breakdown in topological charge. The vacuum fills up with vortices. Presumably  $\mathcal{V}(\psi^* \psi)$  contains repulsive forces which eventually stabilize the proliferation of loops (an example is  $\mathcal{V}(\psi^* \psi) = g \psi^* \psi \psi^* \psi$  which corresponds to repulsive delta-function forces between points on vortices).  $\psi$  then acquires a vacuum expectation value,  $\langle \psi \rangle = \psi_0$ . Because the (denominator) Lagrangian of eq. (3.8) is a function of  $\psi^* \psi(x)$ ,  $\psi_0 \exp(i\theta)$  for  $0 < \theta < 2\pi$  are also action minima. Consider a long rectangular Wilson loop of width  $r$  and length  $t$  (fig. 3a). Trying  $\psi = \psi_0$  as a trial solution yields

$$\langle \exp[iq \oint A \cdot dy] \rangle \sim \exp \left[ - \left( \frac{2\pi q}{e} \right)^2 \psi_0^2 \int d^3x B^2(x) \right] . \quad (3.9)$$

$B(x)$  is given in eq. (3.6) and is the magnetic field created by two parallel wires with opposite current flowing through them. The evaluation of eq. (3.9) is a problem of undergraduate electromagnetism [18]:

$$\int d^3x B^2 \sim \left(\frac{1}{4\pi}\right)^2 2t \ln \frac{r}{r_0} .$$

The constant  $r_0$  should be of the order of the vortex width, since thick vortices can partially intersect a Wilson loop and our idealized approximation breaks down. For  $q/e$  greater than 1, we can find screening-type solutions which better minimize the action. Let  $m$  be the nearest integer to  $q/e$  and set  $\Delta q = m - q/e$ . Let

$$\psi(x) = \exp[i\chi(x)] \psi_0 ,$$

$$\chi(x) \equiv m \left\{ \tan^{-1} \left| \frac{x-x_1}{y-y_1} \right| - \tan^{-1} \left| \frac{x-x_2}{y-y_2} \right| \right\} , \tag{3.10}$$

where  $(x_1, y_1)$  and  $(x_2, y_2)$  are the  $(x, y)$  coordinates of the two lines comprising the Wilson loop. In eq. (3.10)  $m$  must be integer-valued so that  $\chi(x)$  is single-valued

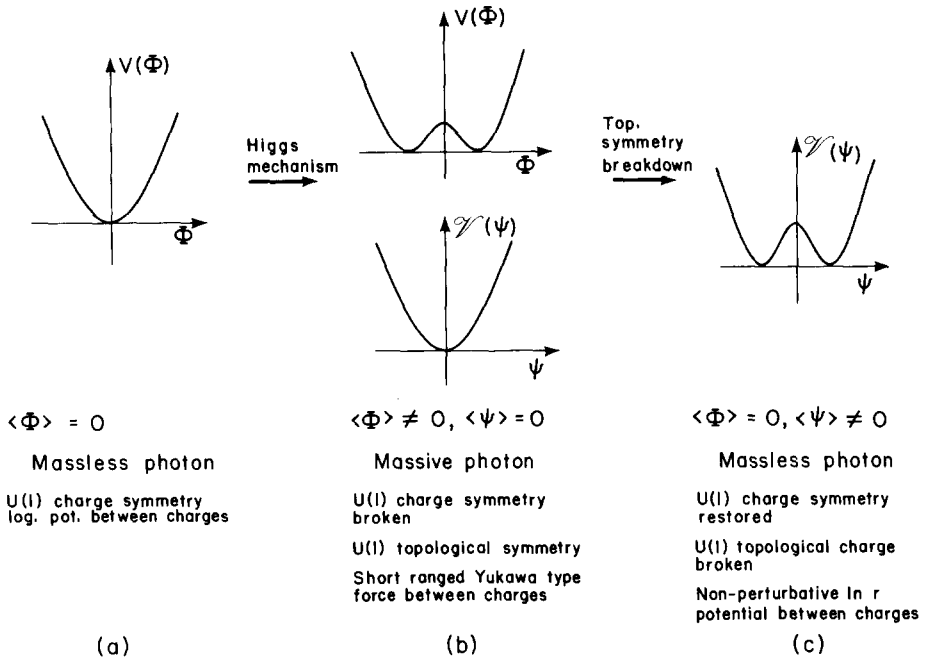


Fig. 4. The phases of the 2 + 1 dimensional Nielsen-Olesen model. (a) The symmetric phase. (b) The Higgs phase. (c) The topological symmetry broken phase.

when circling around  $(x_1, y_1)$  or  $(x_2, y_2)$ . We obtain

$$\langle \exp[iq \oint A \cdot dy] \rangle \sim \exp\left[-\frac{1}{2}(\Delta q)^2 t \psi_0^2 \ln \frac{r}{r_0}\right]. \quad (3.11)$$

We expect eq. (3.10) to be approximately the correct solution for arbitrary  $\mathcal{V}(\psi^* \psi)$ . Only an attractive singular potential could cause the vortices to form neutral bound objects and ruin the picture. Our result is obtained in an almost model-independent manner.

The Wilson loop test is sensitive only to the excess charge. This is the omnipresent periodic (in  $q$ ) screening effect which occurs when the potential is due to topological configurations.

Eq. (3.11) indicates a logarithmic potential between charges because of topological symmetry breakdown. Before  $\psi$  had acquired a vacuum expectation value, there was no logarithmic potential due to the photon because the Higgs mechanism had given the photon a mass. Spontaneous symmetry breakdown has restored a two-dimensional Coulomb-like force.

In addition to this topological non-perturbative  $\ln r$  potential, we expect the perturbative  $\ln r$  potential of the unbroken U(1) gauge theory to be present also: recall property (ii) of the Nielsen-Olesen vortex, that  $\Phi$ , the charged scalar field, must vanish at the vortex. Along the vortex the photon is “massless” in contrast to outside the vortex where it has mass. When topological symmetry breakdown occurs the vacuum is filled with vortices until repulsive forces take over. Presumably this occurs when they begin to overlap. This means that  $\langle \Phi(x) \rangle$  must be zero since vortices occupy all of space-time. The photon must be massless, since  $\langle \Phi(x) \rangle \neq 0$  was the factor contributing to its mass. The original U(1) charge symmetry is restored and another perturbative  $\ln r$  potential due to the photon is expected. The sequel is depicted in fig. 4.

#### 4. $Z(N)$ vortices

We will now consider 't Hooft vortices. We will proceed in a manner similar to sect. 3: first reviewing the soliton solutions and their properties and then performing the Wilson loop calculation.

The topology  $^*$  of an SU( $N$ ) vortex is most easily discussed for  $N = 2$ . For this case, a possible Lagrangian is

$$\begin{aligned} \mathcal{L} = & \frac{1}{4} F_{\mu\nu}^2 + \frac{1}{2} |(\partial_\mu - igA_\mu^a L^a) \Phi_{(1)}|^2 + \frac{1}{2} |(\partial_\mu - igA_\mu^a L^a) \Phi_{(2)}|^2 \\ & + V(\Phi_{(1)}, \Phi_{(2)}), \end{aligned} \quad (4.1)$$

\* See previous footnote.



where

$$V(\Phi_{(1)}, \Phi_{(2)}) = \lambda_1(\Phi_{(1)}^2 - F_{(1)}^2)^2 + \lambda_2(\Phi_{(2)}^2 - F_{(2)}^2)^2 + \lambda_3(\Phi_{(1)} \cdot \Phi_{(2)})^2. \quad (4.2)$$

$\Phi_{(1)}, \Phi_{(2)}$  are two SU(2) triplet Higgs fields,  $L^a$  are the  $3 \times 3$  SU(2) matrix generators, and  $\lambda_1, \lambda_2, \lambda_3, F_{(1)}$ , and  $F_{(2)}$  are constants. Other Higgs potentials will also work.

Because all fields are in the adjoint representation, the symmetry of this model is SU(2)/Z(2) or O(3). The Higgs potential breaks this symmetry completely, so that all three gauge fields acquire mass. The remaining bosonic excitations are also massive. From this point of view, there can be no long-range forces.

As in the Nielsen-Olesen case, there are two types of topological numbers, one related to Higgs fields, the other connected with gauge fields. The solitons have both types.

Consider a static vortex at  $x = y = 0$  (fig. 2). Go far from it and circulate around it. The Higgs fields must take on vacuum expectation values. In going around the circle, these values trace out a curve in the minima, M, of the Higgs potential. Such curves are characterized by  $\pi_1(M)$ . Are there curves in M not deformable to a point for the  $V$  in eq. (4.2)? M is the set of pairs of three-dimensional vectors,  $v_{(1)} \equiv \Phi_{(1)}/F_{(1)}$  and  $v_{(2)} \equiv \Phi_{(2)}/F_{(2)}$  satisfying  $v_{(1)} \cdot v_{(1)} = v_{(2)} \cdot v_{(2)} = 1, v_{(1)} \cdot v_{(2)} = 0$ . If we append the vector  $v_{(3)} = v_{(1)} \times v_{(2)}$ , an orthonormal frame is obtained. Therefore, M is the set of orientations of this frame. Fix a reference frame  $v_{(1)}^0 = (1, 0, 0), v_{(2)}^0 = (0, 1, 0), v_{(3)}^0 = (0, 0, 1)$ . An arbitrary frame is determined by an O(3) rotation of this reference frame. Therefore, M is the set of orthogonal transformations. Characterize the rotation à la Schiff [19] by a vector in the direction of the rotation whose magnitude is the angle of rotation. M becomes isomorphic to the solid three-dimensional sphere with antipodal points identified. This space has curves which cannot be deformed to a point (see fig. 5). However, a path which goes twice along the route of fig. 5 can be deformed to a point (see fig. 6). Physically this is demonstrated in [20]. The fundamental group,  $\pi_1(M)$ , is Z(2), and this character-

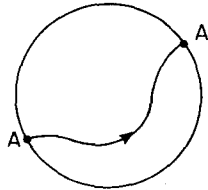


Fig. 5. A non-trivial path. Here is a solid sphere with antipodal points identified. This solid is topologically equivalent to O(3). A closed path is said to be trivial if, via continuous deformations, it can be shrunk to a point. The path shown here begins at A, makes its way through the sphere and ends at A'. The path is closed since A' is the same point as A. Because when A is moved A' must move so as to be opposite A, it is impossible to bring A to A' so as to shrink the path to a point. This path is non-trivial.

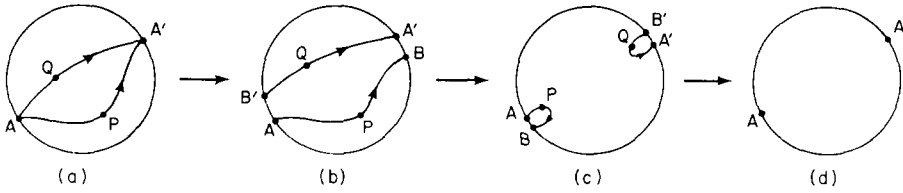


Fig. 6. The product of two non-trivial paths equals a trivial one. Multiplying two paths,  $P_1$  and  $P_2$ , which begin and end at the same point is defined to be the path,  $P_3 = P_2 \cdot P_1$ , formed by first traversing  $P_1$  and then traversing  $P_2$ . (a) Here is a path which begins at A, goes to  $A'$  via P, where it “reappears” at A. It then goes to  $A'$  via Q, hence back to the starting point, A. This path is the product of two “fig. 5” paths. (b) We deform the curve a bit. The new path again starts at A, but instead goes to a point, B, nearby  $A'$ . It “reappears” at  $B'$ , whence it goes to  $A'$  via Q. This closes the path since  $A'$  is identified with A. (c) Move the point B (and hence the point,  $B'$ ) around the sphere until it comes to A. (d) Shrink the two loops to the points A and  $A'$ . Since we have continuously deformed this path to a point, it is trivial.

izes the vortex. The vortices carry a topological charge conserved modulo two. Again do not confuse this  $Z(2)$  with the center of  $SU(2)$ . This is another  $Z(2)$ , topological in nature. Typical non-trivial Higgs configurations are shown in fig. 7.

The topological number associated with the gauge field is similar to the Nielsen-Olesen case. Let  $A_\mu = A_\mu^a L^a$ , where  $L^a$  are the  $3 \times 3$  “angular momentum” matrices.

$$[P \exp \{ ig \oint_x^y A_\mu \cdot dx^\mu \} ]_{\alpha\beta}$$

is the path-ordered product from  $x$  to  $y$ . As  $y$  goes around the circle and back to the starting point,  $x$ , (fig. 2), this matrix traces a curve in  $O(3)$ . Because we are far away, in the region where the Higgs fields are covariantly constant, this matrix is

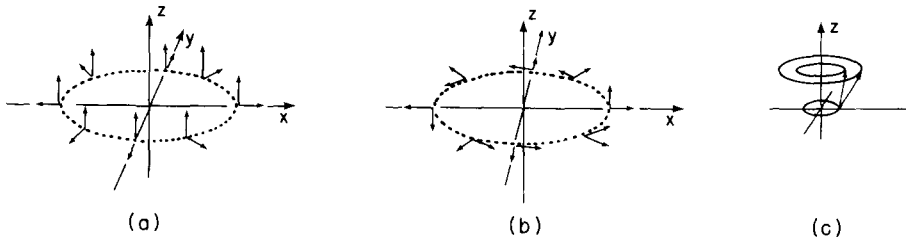


Fig. 7. Non-trivial Higgs configurations. (a) and (b) show the behavior of two non-trivial Higgs configurations far from the vortex. They carry the topological  $Z(2)$  charge. (c) shows what happens near the vortex: the Higgs fields become almost parallel but still rotate when going around the vortex. At the center they become exactly parallel.

precisely the  $O(3)$  rotation that takes the Higgs frame at  $x$  to the Higgs frame at  $y$ . Again a closed curve in  $O(3)$  is obtained which begins and ends at the identity element. Again  $Z(2)$  characterizes the topological charge.

Letting  $A_\mu = \frac{1}{2} \tau^a A_\mu^a$  with  $\frac{1}{2} \tau^a$ , the Pauli matrices:

$$[\mathbb{P} \exp(ig \oint A_\mu \cdot dx^\mu)]_{\alpha\beta} = -\delta_{\alpha\beta}, \quad (4.3)$$

for a gauge field vortex. The line integral is around a circle containing the vortex.

The important local properties of the vertex are:

(i)  $\Phi_{(1)}$  and  $\Phi_{(2)}$  become parallel at the vortex (fig. 7c) so that the frame becomes ill-defined [21];

(ii) for point-like vortices and  $A_\mu^l$  not varying rapidly in  $SU(2)$  space (as the vortex is approached, we expect  $A_\mu^l$  to go to zero rapidly but do not expect the color direction to vary rapidly), there is a delta-function-like contribution to the flux in the direction of  $SU(2)$  space parallel to  $\Phi_{(1)}$  and  $\Phi_{(2)}$ :

$$\hat{\Phi}_{(1)}^l \cdot F_{12}^l(x, y) = \hat{\Phi}_{(2)}^l \cdot F_{12}^l(x, y) \sim \frac{2\pi}{g} \delta(x) \delta(y). \quad (4.4)$$

For  $SU(N)$  we expect the following to be true.

(a) The Higgs potential is chosen to break the symmetry completely. The Higgs fields are in the adjoint representation. For  $SU(3)$  two Higgs fields are sufficient\*. Let  $\Phi_{(i)}^0$  be the vacuum expectation value of the  $i$ th Higgs field. Let  $v_{(i)} = L^l \Phi_{(i)}^0 / F_{(i)}$ , where  $L^l$  are the Lie algebra adjoint matrices of  $SU(N)$ . The symmetry is completely broken if for any  $\eta^l$ ,  $[\eta^l \cdot L^l, v_{(i)}] \neq 0$  for some  $i$ .

(b) Vortices are characterized by  $\pi_1(SU(N)/Z(N)) \approx Z(N)$ .

(c) Define  $v_{(i)}(x) = \hat{\Phi}_{(i)}^l(x) L^l$ . The "hat" over  $\Phi_{(i)}$  indicates that it is normalized to 1, so that

$$\hat{\Phi}_{(i)}^l(x) = \Phi_{(i)}^l(x) / [\Phi_{(i)}^l(x) \Phi_{(i)}^l(x)]^{1/2}.$$

Then at the vortex, there is at least one  $\eta^l$  such that  $[\eta^l \cdot L^l, v_{(i)}(x)] = 0$  for all  $i$ , where  $x$  is the location of the vortex. Let  $\mathcal{F}$  be the set of vectors,  $\eta$ , such that  $[\eta^l \cdot L^l, v_{(i)}(x)] = 0$  for all  $i$ . Then  $a\eta_1 + b\eta_2$  is in  $\mathcal{F}$  and  $[\eta_1, \eta_2]$  is in  $\mathcal{F}$ , if  $\eta_1$  and  $\eta_2$  are in  $\mathcal{F}$ . The latter is true because

$$[[\eta_1, \eta_2], v_{(i)}(x)] = -[[\eta_2, v_{(i)}(x)], \eta_1] - [[v_{(i)}(x), \eta_1], \eta_2] = 0.$$

The set of matrices  $\eta^l \cdot L^l$  for  $\eta \in \mathcal{F}$  forms a Lie subalgebra which generates a subgroup,  $H$ , of  $SU(N)$ .

(d) We expect that only one  $\eta^l$  occurs, so that  $\mathcal{F}$  is one-dimensional and  $H \simeq U(1)$  (in  $SU(3)$ , for example, the two Higgs fields might point in the  $\frac{1}{2}\tau^2$  and  $\frac{1}{2}\tau^3$  directions so that only one  $\eta$  occurs and  $\eta^l = \delta^{l8}$ ). Furthermore  $\eta^l \cdot \frac{1}{2}\lambda^l$  has  $N-1$  eigenvectors with eigenvalue  $1/N$  and one eigenvector of eigenvalue  $(N-1)/N$  when  $\eta^l \cdot \eta^l = 2(N-1)/N$  and  $\frac{1}{2}\lambda^l$  are the fundamental matrix representation normalized

\* We thank Joe Polchinski for this point. Two Higgs fields suffice for all  $SU(N)$ ,  $N \geq 2$ .

so that  $\text{tr} \frac{1}{2} \lambda^l \frac{1}{2} \lambda^m = \frac{1}{2} \delta^{lm}$ . For SU(3) we believe the set of  $\eta^l$  satisfying this condition is of the form

$$\eta \cdot \frac{1}{2} \lambda = \exp(i\alpha \cdot \frac{1}{2}) 2\sqrt{\frac{1}{3}} \frac{1}{2} \lambda^8 \exp(-i\alpha \cdot \frac{1}{2} \lambda),$$

for eight vectors,  $\alpha$ .

(e) There is a delta-function-like contribution to the flux of the form

$$\eta^l \cdot F_{12}^l(x) \sim \frac{2\pi}{g} \delta(x) \delta(y), \quad (4.5)$$

when  $\eta^l$  is normalized so that  $\eta^l \cdot \eta^l = N/2(N-1)$ .

(f) When more than one  $\eta^l$  occurs, the flux might rapidly fluctuate along the vortex in different color directions. Such fluctuations may be the color zitterbewegung phenomenon [22] associated with particles carrying an internal symmetry. There is the speculative possibility that the particles associated with such trajectories are non-Abelian ones and the fields related to them form a representation of H. This is *highly* conjectural and probably impossible to prove. This is a quantum mechanical effect. In this way a dual gauge group may be generated [23]. See also ref. [21].

(g) That the flux must be in the subgroup, H, is reasonable physically. At the vortex gluonic excitations associated with H are massless, because the symmetry, H, is restored. Outside, these same fields are massive. This is a sort of tubular bag-like mass-confinement mechanism of field strengths. They are restricted to the massless regions of space, i.e., the vortices. The important point is that vortices carry tubes of magnetic flux in the group, H.

For the rest of this paper, we will restrict ourselves to the relevant case of SU(3). Let us repeat the calculation of the sect. 3 for SU(3). Idealize to zero-width vortices. The topological charge is conserved mod. 3. Thus, there are two non-trivial types of vortices. One is characterized by

$$\text{tr} [\text{P exp}(ig \oint \mathbb{A} \cdot dl)] = 3 \exp(\frac{2}{3}\pi i),$$

the other by

$$\text{tr} [\text{P exp}(ig \oint \mathbb{A} \cdot dl)] = 3 \exp(-\frac{2}{3}\pi i)$$

(we are now dotting  $A_\mu^l$  into the fundamental representation,  $\frac{1}{2} \lambda^l$ , the  $3 \times 3$  matrices). The path in the path-ordered product is to be taken around the vortex. The 3's in the above equations are trace factors and get replaced by  $N$  for SU( $N$ ). Again, vortices will trace out particle trajectories in three-dimensional Euclidean space. If we assign orientations, then oppositely oriented vortices carry opposite units of flux and may be regarded as antiparticles. The vacuum will be a gas of them. If they have a positive mass they will be small and sparsely located. If they have a negative mass they will fill up the vacuum. The calculation of the Wilson loop in the

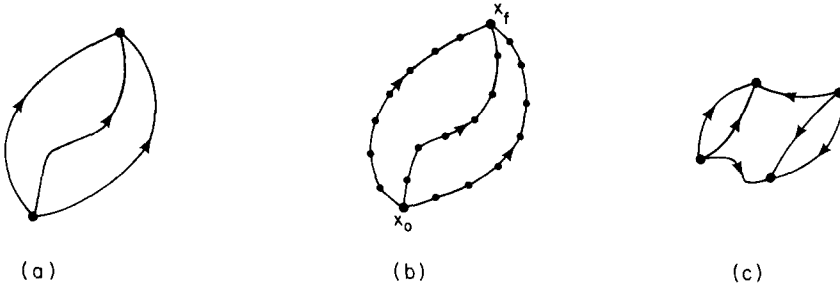


Fig. 8. Triplets. (a) is the simplest triplet and (b) is its macromolecule approximation. (c) is a more complicated structure.

presence of these vortices proceeds as in sect. 3, except that  $q = \frac{1}{3}e$ . The result is

$$\langle \text{tr} [\mathcal{P} \exp(i\oint \mathbf{A} \cdot d\mathbf{x})] \rangle \sim 3 \frac{\iint \mathcal{D}\psi \exp\{-\int d^3x [ |(\partial_\mu - \frac{2}{3}\pi i B_\mu)\psi|^2 + m^2 \psi^* \psi + \mathcal{V}(\psi^* \psi) ]\}}{[\text{same as numerator with } B_\mu = 0]} . \quad (4.6)$$

The function,  $B_\mu$ , is given in eq. (3.6). We have allowed for a mass and for interactions. In principle, these are determined from the original Lagrangian (eq. (4.1) for SU(2)). Again semiclassical methods and/or local field theory soliton methods [13, 21] should be helpful in this respect. The only input in eq. (4.6) is the commutation relations of vortex fields and Wilson loops, which in Euclidean formulation become linking numbers. The factor of 3 is due to color.

The solutions and conclusions are the same as in the Nielsen-Olesen case. There is no long-range potential unless topological symmetry breakdown takes place, in which case the potential is a logarithm. We conclude that *topological symmetry breakdown of  $Z(N)$  vortex loops is not enough to give a linearly confining potential.*

The above calculation considered only closed loops and did not allow for the possibility that three flux tubes could annihilate. Charge conservation, being modulo three, permits such events. Whether it actually happens is a question which can only be answered by finding the effective soliton Lagrangian. It may be that these events occur with zero probability. This question must be answered by doing the proper analysis of the original Lagrangian (the SU(3) analog of eq. (4.1)). No arbitrariness is involved. Let us redo the Wilson loop calculation making the *ad hoc* assumption that three vortices can annihilate. Configurations such as fig. 8a as well as more complicated ones (fig. 8c) are allowed. We shall call these configurations triplets. As indicated by 't Hooft they may be generated by adding a term,  $\lambda_0(\psi^3 + \psi^{*3})$ , to

the Lagrangian of eq. (4.6) \*:

$$\mathcal{L} = |\partial_\mu \psi|^2 + m^2 \psi^* \psi + \mathcal{V}(\psi^* \psi) + \frac{\lambda_0}{3!} (\psi^3 + \psi^{*3}). \quad (4.7)$$

This is seen by doing perturbation theory in  $\lambda_0$ : zeroth order in  $\lambda_0$  is equivalent to our previous gas of interacting loops. Second order (there are no first-order terms) in  $\lambda_0$  produces the configurations of fig. 8a. Higher-order terms yield a gas of those fig. 8c as well as more complicated ones. Consider the second-order term from the macromolecule point of view. Neglect interactions (i.e., set  $\mathcal{V}(\psi^* \psi) = 0$ ) for simplicity.

$$\begin{aligned} \mathcal{Z} &= \mathcal{N} \iint \mathcal{D}\psi \mathcal{D}\psi^* \exp \left[ -\int (\partial_\mu \psi \partial^\mu \psi^* + m_0^2 \psi^* \psi + \frac{\lambda_0}{3!} (\psi^3 + \psi^{*3})) \right] \\ &= \mathcal{Z}_0 \left[ 1 + \frac{\lambda_0^2}{3!} \int d^3x_0 \int d^3x_f \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \int_0^\infty d\tau_3 \iint_{\substack{x_{(1)}(0)=x_0 \\ x_{(1)}(\tau_1)=x_f}} \mathcal{D}x_{(1)} \right. \\ &\quad \times \iint_{\substack{x_{(2)}(0)=x_0 \\ x_{(2)}(\tau_2)=x_f}} \mathcal{D}x_{(2)} \iint_{\substack{x_{(3)}(0)=x_0 \\ x_{(3)}(\tau_3)=x_f}} \mathcal{D}x_{(3)} \\ &\quad \times \exp \left\{ -\int_0^{\tau_1} [\frac{1}{4} \dot{x}_{(1)}^2 + m_0^2] - \int_0^{\tau_2} [\frac{1}{4} \dot{x}_{(2)}^2 + m_0^2] - \int_0^{\tau_3} [\frac{1}{4} \dot{x}_{(3)}^2 + m_0^2] \right\} + O(\lambda_0^4) \\ &\approx \mathcal{Z}_0 \left[ 1 + \frac{\lambda_0^2}{3! \epsilon^3} \left(\frac{1}{6}\right)^3 \int d^3x_0 \int d^3x_f Z_p + O(\lambda_0^4) \right], \end{aligned} \quad (4.8)$$

where

$$\begin{aligned} Z_p &= \epsilon^9 \sum_{N_1=1}^\infty \sum_{N_2=1}^\infty \sum_{N_3=1}^\infty \left[ \prod_{i_1=1}^{N_1-1} \int d^3x_{i_1}^{(1)} \right] \left[ \prod_{i_1=1}^{N_1} \frac{\delta(|x_{i_1}^{(1)} - x_{i_1-1}^{(1)}| - \epsilon)}{4\pi\epsilon^2} \right] \\ &\quad \times \left[ \prod_{i_2=1}^{N_2-1} \int d^3x_{i_2}^{(2)} \right] \left[ \prod_{i_2=1}^{N_2} \frac{\delta(|x_{i_2}^{(2)} - x_{i_2-1}^{(2)}| - \epsilon)}{4\pi\epsilon^2} \right] \\ &\quad \times \left[ \prod_{i_3=1}^{N_3-1} \int d^3x_{i_3}^{(3)} \right] \left[ \prod_{i_3=1}^{N_3} \frac{\delta(|x_{i_3}^{(3)} - x_{i_3-1}^{(3)}| - \epsilon)}{4\pi\epsilon^2} \right] \\ &\quad \times \exp \left\{ -\sum_{i_1=1}^{N_1} \beta\mu - \sum_{i_2=1}^{N_2} \beta\mu - \sum_{i_3=1}^{N_3} \beta\mu \right\}. \end{aligned} \quad (4.9)$$

\* Many other terms can be added which also generate triplets. They are non-local interaction terms. They are permitted, of course, since there is no reason to expect the effective soliton Lagrangian to be a local field theory.

*Comments:* (a)  $\mathcal{Z}_0$  is the previously discussed grand partition function for a gas of non-interacting 't Hooft vortex loops.  $Z_p$  is the partition function for three macromolecules which all begin at  $x_0$  and all end at  $x_f$  (fig. 8b). The endpoints,  $x_0$  and  $x_f$ , are arbitrarily located in three-dimensional Euclidean space, hence the integrals,  $\int d^3x_0 \int d^3x_f$ . There is also an arbitrary number of atoms,  $N_1, N_2$  and  $N_3$ , for each macromolecule;  $x_s^{(j)}$  is the position of the  $s$ th atom of the  $j$ th macromolecule. The chemical potential is  $\mu = \epsilon^2 m_0^2 / \beta$ , where  $\beta$  is the inverse temperature and  $\epsilon$  is the length of a bond. The latter is the cutoff parameter in our segmented line approach to a continuous curve.

(b) The three macromolecules in eq. (4.9) are 't Hooft vortices. Reintroducing  $\mathcal{V}(\psi^* \psi)$ , they undergo the same monomer or bond interactions as loops. Each carries  $2\pi/g$  units of flux from  $x_0$  to  $x_f$ .

(c) Because  $\mathcal{Z}_0$  multiplies  $Z_p$ , the second term,  $\mathcal{Z}_0 Z_p$ , is a system of an arbitrary number of closed loops and one triplet configuration. Let

$$\lambda_p = \frac{\lambda_0^2}{3! \epsilon^3} \left(\frac{1}{6}\right)^3 \equiv (\lambda_R^m)^2 \tag{4.10}$$

be a renormalized triplet activity (the subscript p stands for pair since there is a pair of vertices in figs. 8a and 8b). Higher order terms in  $\lambda_p$  will generate multiple triplet configurations and will lead to  $\mathcal{Z}_0(\sum_{M=0}^\infty (1/M!) \lambda_p^M Z_p^M)$ , a grand partition function for a system of closed loops and triplets. The combinatorial factors in Feynman rules precisely give the  $1/M!$  factor necessary for a grand partition function (vacuum bubbles exponentiate).

Let us redo the Wilson loop integral test, allowing for triplets and more complicated configurations. Both loops and triplets lead to phase factors dependent on linking number. Fig. 9 illustrates some possibilities.

Trouble arises in trying to repeat the calculation of sect. 3. Eq. (3.4) is no longer valid. It works only for closed loops and not for those of fig. 8. Fortunately, eqs.

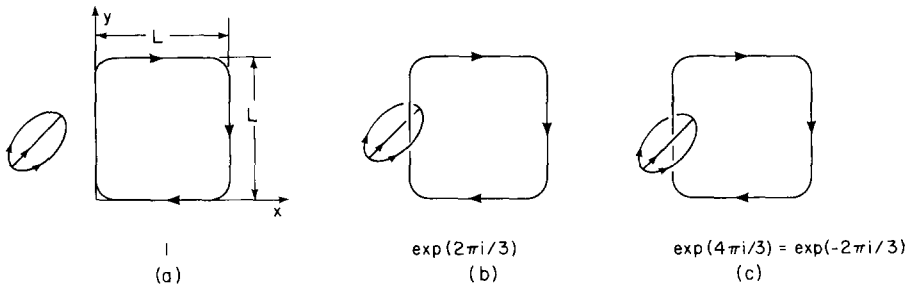


Fig. 9. Linking number of a triplet with a Wilson loop. (a) The Wilson loop. It has dimensions  $L \times L$  and sits in the  $x$ - $y$  plane. The  $z$ -axis comes out of the paper. The nearby triplet does not link with the Wilson loop so  $n = 0$ . (b) A linking configuration yielding the phase factor  $\exp(2\pi i/3)$  and (c) a linking configuration yielding the phase factor  $\exp(-2\pi i/3)$ .

(3.4) and (3.5) are not the only ones for linking number. Suppose  $B_\mu$  is replaced by  $B_\mu + \partial_\mu \chi$ , where  $\chi$  is an arbitrary smooth function. Then, according to eq. (3.5),

$$\begin{aligned} n &= \sum_i \int_{C_i} B_\mu(x) dx^\mu \rightarrow \sum_i \int_{C_i} (B_\mu + \partial_\mu \chi) dx^\mu \\ &= \sum_i \int_{C_i} B_\mu(x) dx^\mu . \end{aligned}$$

The effect of  $\partial_\mu \chi$  disappears because of Stokes theorem; there is a kind of “gauge invariance” in defining  $B_\mu$ . Does this arbitrariness affect the conclusions of sect. 3? The answer is no, as seen from eq. (3.8). A change in  $B_\mu$  by  $\partial_\mu \chi$  can be absorbed in the solution for  $\psi$  by multiplying  $\psi(x)$  by  $\exp[i(2\pi q/e) \chi(x)]$ . This leads to the same action and the same  $\ln r$  behavior of the potential. The Lagrangian of eq. (3.8) has a global U(1) invariance which allows a redefinition of  $B_\mu$  to be absorbed in a redefinition of  $\psi$ . Previously, the 't Hooft vortex Lagrangian (without the  $\lambda_0(\psi^3 + \psi^{*3})$  term) had this U(1) invariance also. The charges  $\frac{2}{3}\pi$  and  $-\frac{2}{3}\pi$  were absolutely conserved. Such a system looked like ordinary charge. Only when  $\psi^3$  type terms are added can one “see” charge conserved modulo three. This is why the conclusions of the 't Hooft model were similar to the Nielsen-Olesen case. Now that triplets are present we expect different conclusions. We can use the “gauge invariance” to define a  $B_\mu$  which works. Let S be any two-dimensional surface which spans the Wilson loop; the boundary of S is the Wilson loop. For simplicity take S to be a surface of minimal area. Let

$$B_\mu(x) = \int_S \delta^3(x - y) dS^\mu(y) . \tag{4.11}$$

Here  $dS^\mu$  is the surface element directed normal to the surface (the sign of the normal is determined by the orientation of the Wilson loop). As a clarifying example, take the Wilson loop in fig. 9a:

$$B_\mu(x) = -\delta_{\mu 3} \delta(z) \theta(L - x) \theta(y) \theta(L - y) . \tag{4.12}$$

If a curve,  $x(\tau)$ , pierces the surface then  $\int d\tau \dot{x}_\mu(t) B^\mu(x(t))$  is plus or minus one depending on the piercing direction. Roughly speaking,  $B_\mu(x)$  (of eq. (4.11)) points in the direction normal to the surface, acts like a delta function in this direction, and vanishes away from the surface. This  $B_\mu(x)$  is obtainable from the old one by a gauge transformation. The old  $B_\mu(x)$  (of eq. (3.6)) is in the “Lorentz gauge”,  $\partial^\mu B_\mu = 0$ . The new one is in a “surface axial gauge”. Both  $B_\mu$ 's have the same curl. The new  $B_\mu$ , however, can handle triplets. This follows from the above discussion when one does perturbation theory in  $\lambda_0$  and returns to the macromolecule analog gas.



We proceed as in sect. 3. The result is

$$\langle \text{tr} [\text{P exp}(ig \oint A \cdot dI)] \rangle \tag{4.13}$$

$$= \frac{\iint \mathcal{D}\psi \mathcal{D}\psi^* \exp\{-\int [ |(\partial_\mu - i\frac{2}{3}\pi B_\mu)\psi|^2 + m_0^2 \psi^* \psi + \mathcal{V}(\psi^3, \psi^{*3}, \psi^* \psi) ]\}}{\text{[same term as numerator with } B_\mu = 0 \text{]}}$$

Because of the singular nature of  $B_\mu$ , the action in the numerator is infinite unless  $\psi$  vanishes on  $S$ . The solutions are as follows.

(a) When  $m_0^2 > 0$  and  $\langle \psi \rangle = 0$ ,  $\psi = 0$  is the solution and there is no confinement due to 't Hooft vortices.

(b) When  $m_0^2 < 0$  and  $\langle \psi \rangle \neq 0$ , the equations of motion must be solved in the presence of  $B_\mu(x)$ . The precise nature of the solution is described below in paragraph (vii). We find that  $\psi$  has non-vacuum expectation values near  $S$ , that the action goes like the area, and that the potential grows with  $r$ . *Spontaneous symmetry breakdown with triplets present gives linear confinement.*

We now make some observations.

(i) The effect of choosing a general gauge for  $B_\mu$  is as follows: suppose the  $B_\mu$  of eq. (4.11) is replaced by  $B_\mu + \partial_\mu \chi$ . When expanding eq. (4.13), the macromolecules in a triplet, going from  $x_0$  to  $x_f$ , get multiplied by  $\exp[3i\chi(x_f) - 3i\chi(x_0)]$ . This transformation is innocuous only if  $\chi(x_0)$  and  $\chi(x_f)$  are multiples of  $\frac{2}{3}\pi$ . Because there is a gas of triplets, whose position may be anywhere, general gauges are not allowed. The arbitrariness in the triplet's location constrains  $\chi(x)$ . A gauge transformation causes triplets to be multiplied by unwanted phase factors and ruins the Wilson loop calculation. A singular surface gauge must be chosen.

(ii) What happens if a non-minimal surface is chosen? Does the action go as the area of this surface? Consistency demands that the physics be independent of  $S$ . Suppose another non-minimal surface,  $S'$ , is chosen. The minimal surface,  $S$ , and the non-minimal surface,  $S'$ , form a closed surface (fig. 10). Let  $V$  be the enclosed

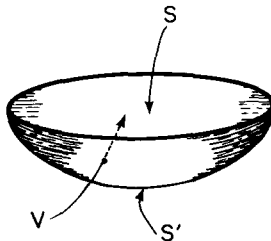


Fig. 10. Spanning surfaces. The dark line is the Wilson loop. The non-minimal surface,  $S'$ , is the "cup-like" surface below the loop. The minimal surface,  $S$ , forms a "lid". Together they enclosed the volume,  $V$ .

volume. Redefine  $\psi$  by

$$\psi(x) \rightarrow \psi(x) \exp[i \frac{2}{3}\pi \chi_V(x)] , \quad (4.14)$$

where  $\chi_V(x)$  is the characteristic function for  $V$ , that is,  $\chi_V(x) = 1$  if  $x$  is in  $V$  and  $\chi_V(x) = 0$  if  $x$  is outside  $V$ . Plugging into eq. (4.13) with  $B_\mu$  given by eq. (4.11) one sees that

$$(\partial_\mu \psi - i \frac{2}{3}\pi B_\mu^{S'} \psi) \rightarrow (\partial_\mu \psi - i \frac{2}{3}\pi B_\mu^S \psi) . \quad (4.15)$$

The surface may be moved around by doing a step-function gauge transformation where steps occur in multiples of  $\frac{2}{3}\pi$ . According to (i), such a gauge transformation is allowed. In eq. (4.15)  $B_\mu^{S'}$  is the  $B_\mu$  of eq. (4.11) for the surface,  $S'$ , and  $B_\mu^S$  is the  $B_\mu$  for the surface,  $S$ . Thus, for a non-minimal surface, the solution is the one for the minimal surface multiplied by a step-function phase factor and leads to the same action. The Wilson loop action again goes as the minimal spanning surface area.

(iii) The solution is periodic in  $q = 2\pi n$ : when a higher-dimensional representation is used so that the effective charge is  $q = 2\pi n$ , ‘‘screening’’ occurs and the action no longer goes like the area. The Lagrangian

$$\mathcal{L} = |(\partial_\mu - iqB_\mu)\psi|^2 + m^2 \psi^* \psi + \mathcal{V}(\psi^* \psi, \psi^3, \psi^{*3}) , \quad (4.16)$$

for  $q = 2\pi n$  has trivial solutions where the phase of  $\psi$  jumps by  $2\pi n$  across the surface,  $S$ . The singularity in taking the derivative of this phase cancels the singularity in  $B_\mu$ . Since the phase of  $\psi$  is defined modulo  $2\pi$ , there is no mismatch of phases in  $\psi$  when going around the line of a Wilson loop; there are no global difficulties with this solution. For  $q = 2\pi n \pm \frac{2}{3}\pi$ , a similar procedure yields solutions whose action is the same as  $q = \pm \frac{2}{3}\pi$ . Quarks will be confined but gluons will be screened. Going back to the original macromolecule partition function, one sees trivially that the Wilson loop is 1 for  $q = 2\pi n$ . However, it is non-trivial that classical solutions reproduce this phenomenon since the saddle point is an approximation. This gives confidence to our methods.

(iv) The solution is virtually independent of  $\mathcal{V}(\psi^* \psi, \psi^3 \psi^{*3})$  although a strong potential between vortices may cause them to form dipole-like objects and ruin confinement, an unlikely possibility we feel. Thus confinement is a general phenomenon (almost) independent of the forces between vortices. However,  $m^2$  being less than (or equal to) zero must be the reason that  $\langle \psi \rangle \neq 0$  because of our reliance on the analogy with a gas of loops. By  $m$  we mean the physical effective mass which includes the energy per unit length as well as entropy embedding effects. There are other ways in which  $m^2 > 0$  but  $\langle \psi \rangle \neq 0$  (see the potential of fig. 11). These potentials result in a liquid-gas-type phase transition discussed by Langer [24] and Coleman [25]. Bubbles of true vacuum form rather than a dense spaghetti vacuum. The phase transition occurs *via* barrier penetration instead of vacuum instability.

(v) The difference between having and not having triplets is the difference between having and not having a Goldstone phenomenon with a broken invariance. Without triplets the Lagrangian of eq. (4.7) (with  $\lambda_0 = 0$ ) has a global  $U(1)$  symme-

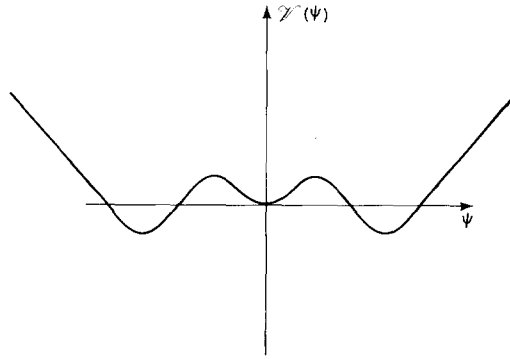


Fig. 11. A “bad” topological symmetry breaking potential.

try, because, as we have noted, topological charge, unable to be created or destroyed, behaves like ordinary charge. Ordinary charge is associated with a  $U(1)$  invariance. When spontaneous symmetry breakdown takes place, a Goldstone boson occurs. It is this massless particle which creates long-ranged forces which ruin confinement and lead to only a logarithmic potential, although we cannot explicitly demonstrate this. Contrast this to when triplets occur. Charges in threes are created and destroyed; the symmetry is  $Z(3)$ , a discrete group. In this case, no Goldstone boson occurs to disrupt the linear confinement. Proving triplets exist will be difficult. We know no obvious way to use semiclassical methods to calculate  $\lambda_0$ . It is also easy to overlook such configurations using Bardakci's and Samuel's local field theory formulation [13,21] because of the powers of  $\epsilon$  in eq. (4.9).

(vi) We conjecture that when  $\langle\psi\rangle \neq 0$  the symmetry subgroup,  $H$  (which is probably  $U(1)$ ), is restored. Along the vortices the Higgs fields are ineffective in breaking  $H$ , the gauge fields associated with  $H$  are massless, and the symmetry is present. At the vortex the symmetries associated with the generators commuting with “parallel” Higgs fields are not broken. When  $m^2 < 0$ , the vacuum is filled with vortices, so that, virtually, in every square centimeter of space the symmetry is restored. We conclude that topological symmetry breaking will restore at least a  $U(1)$  subgroup of the original color group. Hence, in addition to the linear confinement there will be a logarithmic potential due to these gluons. This logarithmic potential will be analytic in  $g$  and calculable *via* perturbation theory, whereas the topologically generated linear potential is non-analytic in  $g$ .

(vii) What do our solutions look like? In particular, how do our ideas relate to 't Hooft's and are there any differences?

Let us first reproduce this result that  $\text{tr P}[\exp(ig \oint A \cdot dI)]$  creates a region of vacuum with  $\langle\psi(x)\rangle = \psi_0 \exp(\frac{2}{3}\pi i)$  for  $x$  inside  $C$  (throughout our discussion  $\psi_0$  is the vacuum expectation value of  $\psi$  and  $C$  is a closed loop contained in a time slice of three-dimensional Euclidean space). In a Euclidean formulation this is seen as

follows: consider the propagator of two Wilson loops

$$\langle \text{tr}[\text{P exp}(-ig \oint_{C'} \mathbf{A} \cdot d\mathbf{l})] \text{tr}[\text{P exp}(ig \oint_C \mathbf{A} \cdot d\mathbf{l})] \rangle, \\ -\langle \text{tr}[\text{P exp}(-ig \oint_{C'} \mathbf{A} \cdot d\mathbf{l})] \rangle \langle \text{tr}[\text{P exp}(ig \oint_C \mathbf{A} \cdot d\mathbf{l})] \rangle, \tag{4.17}$$

where  $C'$  occurs at a much later time than  $C$  (see fig. 12). To evaluate eq. (4.17), resort to the “gas of loops” analogy used to calculate a single Wilson loop (see eq. (3.8)). Choose  $B_\mu$  (eq. (4.11)) to be a sum of two terms, one resulting from using the minimal surface of  $C$  and one resulting from the minimal surface of  $C'$ . For  $t$  large, the solution to the equations of motion is approximately the sum of the solutions of each Wilson loop. When the vacuum expectation value of each is subtracted off as in eq. (4.17), the contribution cancels. There are, however, other surfaces which span a pair of Wilson loops, which are not the union of two surfaces, one for  $C$  and one for  $C'$  (see fig. 12). They look like “hour glasses”. They occur when  $\psi(x) \sim \psi_0 \exp(\frac{2}{3}\pi i)$  inside the “hour glass”, in which case, the singularity in  $B_\mu$  on  $C$  and  $C'$  is cancelled and reappears on the “hour glass” surface. The contributions from these do not cancel in eq. (4.17). For the new  $\psi$  the action goes like the hour glass’s surface area. Classically, the hour glass will try to be small-necked. This, of course, is the instability of a classical closed string. Quantum mechanically, there is a sum over all surfaces, each weighted by its surface area (the Nambu action). Eq. (4.17) will result in the propagator of two closed strings. These new solutions yield a  $\psi$  with the extra phase factor,  $\exp(\frac{2}{3}\pi i)$ , inside. This coincides with 't Hooft’s conclusion; the physical interpretation of this process is that  $\text{tr}[\text{P exp}(ig \oint_C \mathbf{A} \cdot d\mathbf{l})]$  produces a region of  $\psi_0 \exp(\frac{2}{3}\pi i)$  vacuum which propagates until  $\text{tr}[\text{P exp}(-ig \oint_{C'} \mathbf{A} \cdot d\mathbf{l})]$  destroys it. The Wilson loop operator does, indeed, produce regions of topological  $Z(3)$  vacuum.

Now let us deal with a quark-antiquark system or equivalently put a Wilson loop

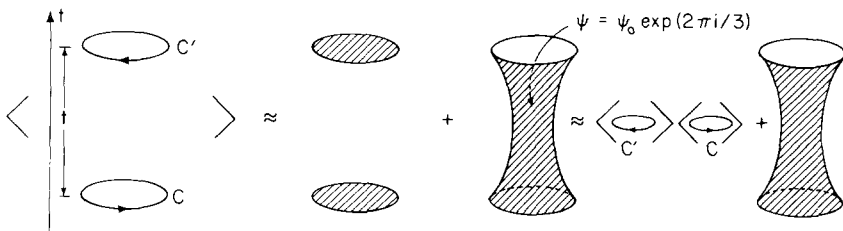


Fig. 12. The propagator of two Wilson loops,  $C$  and  $C'$ . There are two types of spanning surfaces giving contributions; the ones where each Wilson loop annihilates into the vacuum and the “hour-glass” ones.

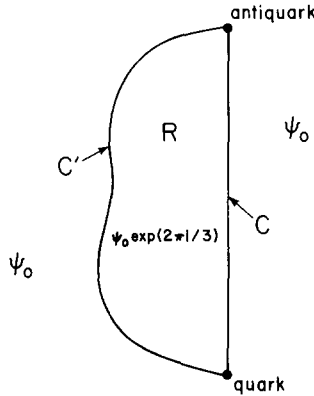


Fig. 13. A time slice of a Wilson loop. At an instant in time, the non-minimal gauge surface,  $S'$ , and the minimal surface,  $S$ , will look like the curves  $C'$  and  $C$ . The classical solution has the phase factor  $\exp(\frac{2}{3}\pi i)$  inside the region,  $R$ , enclosed by  $C$  and  $C'$ .

in the system. Choose a non-minimal surface,  $S'$ , as the surface in the surface axial gauge. By looking at a time slice, we can relate our solutions to a Hamiltonian picture. Fig. 13 shows, in this time slice, the curve  $C'$ , which is the intersection of  $S'$  with this slice, and the curve  $C$ , which is the intersection of the minimal surface,  $S$ , with this slice.  $C'$  and  $C$  enclose a region,  $R$ . The classical solution has  $\psi(x) = \psi_0$  far from  $R$ ,  $\psi(x) = \psi_0 \exp(\frac{2}{3}\pi i)$  well inside  $R$ . Along  $C'$  the phase of  $\psi(x)$  jumps discontinuously by  $\frac{2}{3}\pi$ . The singularity in  $\partial_\mu \psi(x)$  thereby cancels the singularity in  $B_\mu(x)$ . As  $x$  goes from inside  $R$ , across  $C$ , to a point far from  $R$ ,  $\psi(x)$  must return to  $\psi_0$ . Thus, in a region surrounding  $C$ ,  $\psi$  undergoes a phase change by  $\frac{2}{3}\pi$ . The solution must have the boundary conditions that  $\psi(x) = \psi_0 \exp(\frac{2}{3}\pi i)$  to the left of  $C$  and  $\psi(x) = \psi_0$  to the right. The solution which satisfies these constraints will be the soliton discussed by 't Hooft [1]. The coefficient of  $r$  in the confining potential is the mass of the Bloch wall soliton. We have reproduced 't Hooft's conclusion exactly using our formalism. One also sees the very unphysical nature of  $S'$ .



Fig. 14. Quark trajectories in a baryon. (a) A baryon consisting of three quarks is produced at  $a$  and destroyed at  $b$ . In between the three quarks travel along paths  $C_1$ ,  $C_2$  and  $C_3$ . (b) A singular gauge surface. It consists of three disks, each of which is bounded by a quark trajectory and the line from  $a$  to  $b$ .

(viii) The confinement criterion for baryons is different from mesons [22]. What replaces the Wilson loop is (see fig. 14a)

$$\begin{aligned} & \epsilon_{\alpha\beta\lambda}\epsilon_{\alpha'\beta'\lambda'} [P \exp(ig \int_{C_1}^b \mathbb{A} \cdot dI)]_{\alpha\alpha'} [P \exp(ig \int_{C_2}^b \mathbb{A} \cdot dI)]_{\beta\beta'} \\ & \times [P \exp(ig \int_{C_3}^b \mathbb{A} \cdot dI)]_{\lambda\lambda'} . \end{aligned} \quad (4.18)$$

Take  $C_1$  far from  $C_2$  and  $C_3$ . If the action decreases exponentially with area then confinement in baryons occurs. Fig. 14a resembles the “dual” of a triplet. Similarly, an arbitrary gauge for  $B_\mu$  is not possible. One, therefore, might think that even without triplets topological symmetry breakdown would confine quarks in baryons. This is not true, though: first choose a gauge which is singular on three surfaces such as in fig. 14b. A gauge transformation,  $B_\mu \rightarrow B_\mu + \partial_\mu \chi$ , can be performed as long as  $\chi(a) = (2\pi/3)n$  and  $\chi(b) = 2\pi n/3$  (a and b are the endpoints in fig. 14). Hence, a gauge can be chosen for which  $B_\mu$  is smoothed out. If a and b are far apart, it can be made to look like the Lorentz gauge in the region far from both a and b. This will lead to only a logarithmic potential. *Triplets are needed for baryon confinement also.* Concerning the restrictive effect on gauge choice, the difference between baryons and triplets is that the latter form a gas. Endpoints of triplets can be located anywhere; given any space-time point there is a configuration in the statistical ensemble with a triplet vertex there. This forces  $\chi$  to be a step function in units of  $\frac{2}{3}\pi$  everywhere. On the other hand a baryon constrains  $\chi$  at only two points, a and b.

In the presence of triplets and topological symmetry breakdown, baryons will be confined. A gauge for  $B_\mu$  singular on surfaces (fig. 14b) must be chosen. One must solve the  $\psi$  equations of motion in the presence of such a  $B_\mu$ .

Assuming constituent quarks are far apart and that the energy goes as the area of the singular surface, we can see how a static baryon looks. In dual string models, there were several speculations: (a) three quarks at the ends of three strings with the other three string endpoints joined (fig. 15a), (b) one quark in the middle of a single string with the other two quarks at the endpoints (fig. 15b), and (c) quarks in a triangular string configuration (fig. 15c). Case (c) cannot occur in our formalism. Case (b) is a special case of (a) when one of the three strings has zero length. When does (a) occur and when does (b) occur? As an example of what happens, constrain the third quark to be equidistant from the other two (fig. 15a). The energy of this configuration is

$$E(d) = m_c(2\sqrt{L^2 + d^2} + y - d) . \quad (4.19)$$

The notation is as in fig. 15a and  $m_c$  is a constant. The point  $d$  is determined by

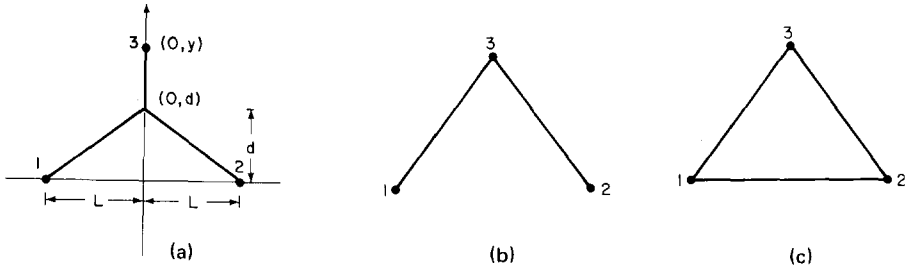


Fig. 15. The shape of baryons. (a) Quarks at the ends of three strings. A time slice of fig. 15b would yield this configuration. (b) A quark in the middle of a string. (c) A triangular quark and string configuration. The numbers 1, 2 and 3 label the quarks. The dark lines are the strings.

setting  $\partial E/\partial d = 0$ , for which we find that for  $d \leq \sqrt{\frac{1}{3}}L$  the third quark sits in the middle (case (b)), whereas for  $d > \sqrt{\frac{1}{3}}L$  three strings form (case (a)). This is reasonable physically: as the third quark moves farther away, energetically it becomes favorable to pull a new string out, rather than stretch two strings.

Because of the different baryon string picture, baryon Regge trajectories might behave differently from meson ones. It may be that at low energies the third quark sits in the middle. This would give similar Regge trajectories and slopes. At higher energies, third quark excitations might form causing baryon trajectories to become different from meson ones.

Let us summarize the key points when topological symmetry breakdown in the 't Hooft vortex charge takes place.

(I) A logarithmic potential is obtained in the absence of triplets and a linear potential is obtained in their presence.

(II) The potential is triality dependent. Representations with integral hypercharge are screened. Representations with fractional hypercharge are confined. The potential is periodic in hypercharge.

(III) Original color symmetries are at least partially restored.

## 5. Monopoles

In this section, we shall show how monopoles arise in the 't Hooft  $Z(N)$  model. The important conclusion will be that, in the presence of triplets, the phase transition from  $\langle \psi \rangle = 0$  to  $\langle \psi \rangle \neq 0$  is a transition from an ensemble of monopole-anti-monopole pairs (magnetic dipoles) to a liberated plasma of monopoles and anti-monopoles. This phase transition might be compared to that of a two-dimensional Coulomb gas [10,26] except that the interaction between monopoles gets changed from a linear one to a non-confining one, most likely a Yukawa,  $(1/r) \exp(-\mu r)$ , potential (the expected Coulomb-like  $1/r$  potential probably gets screened due to plasma effects [7,8]).

Two points need clarifying: in 3 + 1 dimensions for an Abelian theory, we know what a monopole is; we compute  $\int_S B \cdot dA$  over a closed surface, S. If  $2\pi/e$  results, there is a monopole of strength  $2\pi/e$ . First, what is a monopole in 2 + 1 dimensions? Solitons in 3 + 1 dimensions are instantons in 2 + 1 dimensional Euclidean space. The monopoles in the 't Hooft model will be instantons (like Polyakov's, for example [7]). Secondly, what is a monopole in an SU(3) gauge theory? Without Higgs fields, we don't know how to precisely define one, but with Higgs fields we can. A classic example is the 't Hooft-Polyakov monopole [30] in the Georgi-Glashow model. Recall that 't Hooft defined a gauge-invariant  $F_{\mu\nu}$  which, in the presence of the "hedgehog" solution, behaved as a monopole field. The Higgs fields were an essential ingredient in  $F_{\mu\nu}$ . Likewise, we can define an  $F_{\mu\nu}$ , but only when  $H = U(1)$  \* (H, described in sect. 4, is the subgroup in which the flux points). When this happens, there is a vector,  $\eta$  (eq. (4.5)), constructed out of Higgs fields which indicates the color direction of vortex flux. This vector transforms in the adjoint representation under gauge transformations. Hence  $\eta \cdot F_{\mu\nu}$  will be gauge invariant and can be used to measure the flux \*\* . Normalize  $\eta$  so that  $\eta \cdot \eta = N/[2(N - 1)] = \frac{3}{4}$  for SU(3). *Definition:* There is a monopole of  $N$  Dirac units if

$$\int_S B \cdot dA = \frac{2\pi N}{g} \quad , \quad (5.1)$$

where S is a closed surface and  $B^i = \frac{1}{2}\epsilon^{ijk}\eta^j F_{jk}^i$ .

With this definition triplet configurations are monopole-antimonopole instanton pairs. Perform the measurements,  $\int_S B \cdot dA$ , over a sphere enclosing one end of a triplet (fig. 16a). Since each vortex acts like a delta function of flux, we will measure three contributions of flux of  $2\pi/g$  units. The total flux emanating from the endpoint is  $6\pi/g$ . *Triplets are monopole-antimonopole pairs.* Each has three Dirac units. Ordinary vortex loops are not monopoles because  $2\pi/g$  units enter at one point but exit at another (fig. 16b).

Without Higgs fields it is difficult to know what constitutes a monopole. It is important to know whether it can be defined in a pure SU(3) gauge theory. Aesthetically, one would like to do away with the Higgs fields. They are only being used as a crutch.

One might try the following. Take a sphere, S. Break it into small regions,  $R_i$ . Each region,  $R_i$ , has a closed boundary,  $C_i$ . Make measurements [ $\text{tr P exp}(ig \oint_{C_i} \mathbb{A} \cdot dI)$ ]. Define  $1/g$  times the phase to be the "flux". Add up all the fluxes to obtain the total flux. Several problems arise: first,  $\theta$  units of flux is indistinguishable from  $2\pi n + \theta$  units. In treating triplets, one might conclude that the first two vortices contribute  $2\pi/g$  while the last contributes  $-4\pi/g$  so that the total is zero. This pro-

\* For groups, H, larger than U(1), it becomes just as problematic to define a monopole as in a non-Abelian gauge theory without Higgs fields.

\*\* The analog of  $\eta \cdot F_{\mu\nu}$  for the model of ref. [27] is  $\tilde{\Phi} \cdot F_{\mu\nu}$ . This is not the 't Hooft  $F_{\mu\nu}$  but it works equally well as noted by Coleman in ref. [28] and emphasized in ref. [16].



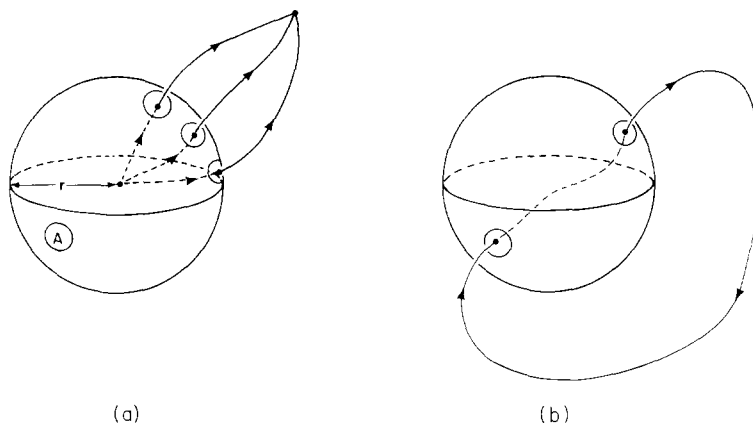


Fig. 16. Monopoles. (a) A sphere of radius  $r$  surrounding the endpoint (monopole) of a triplet. The total magnetic flux emanating is  $6\pi/g$ . The region A is a disk on the sphere. (b) For a vortex the total flux emanating is zero.

cedure would give an incorrect result. Secondly, it is unclear what is being measured because there is no Stoke's theorem for non-Abelian gauge theories <sup>\*</sup>. We are not really measuring the total flux: because non-Abelian flux is not additive like Abelian flux, this procedure will almost always yield a non-zero result. Although gauge invariant, it is useless. Another attempt chooses a vector,  $V^l(x)$ , appropriately normalized. If

$$\frac{1}{2} \int_S V^l \epsilon_{ijk} F_{ij} dS^k \neq 0, \quad (5.2)$$

one might say there is a monopole. For an arbitrary  $V^l$ , this is not gauge invariant so that eq. (5.2) is also meaningless. Fortunately the  $Z(N)$ 't Hooft model has Higgs fields and we are able to circumvent these difficulties. We shall return to this point in sect. 8.

When  $\langle \psi \rangle = 0$  and  $m^2 > 0$ , the three string of a monopole carry energy per unit length (these strings are the analogs of superconductor vortices). This means that a monopole must always be paired very closely to its antimonopole partner. Detection of monopoles is difficult unless the sphere,  $S$ , in eq. (5.1) is minuscule. Dipoles will also have little effect on confinement as the calculations of sect. 4 demonstrated. A dipole's three strings can link with a Wilson loop only when it is near the quark trajectory. This produces only a perimeter effect and a mass renormalization. Next consider what happens when  $m^2 < 0$  and topological spontaneous symmetry breakdown takes place. The chemical potential per atom (in the macromolecule analogy)

<sup>\*</sup> See, however, refs. [29,16].

is negative and vortices of large size are favored. One can see that the monopole-antimonopole constituents in a dipole are liberated by considering a triplet with endpoints far apart. Focus on one of the monopoles. There are three flux strings emanating from it. Unlike the previous case, these strings do not head directly for the partner antimonopole. Instead they are more in arbitrary directions. The partition function (or the Feynman path integral in a particle dynamics description) sums over all these directions. Consider measuring the flux,  $\int_A B \cdot dS$ , flowing through a small disk,  $A$ , of a sphere enclosing the monopole (see fig. 16a). In doing this measurement only a fraction of the time will a string pierce  $A$ . This averaging effect spreads the flux. Symmetry demands we measure  $(6\pi/g) A/4\pi r^2$ , where  $A$  is the area of the disk and  $r$  is the radius of the sphere. In short, averaging over all vortex paths quantum mechanically spreads out the flux so that a “normal” monopole with a radial magnetic field is observed. Since the action of the gauge fields is  $-\frac{1}{4} \int F_{\mu\nu}^2 d^3x$ , we expect  $\star$  that these monopoles will interact very much like ordinary ones, with Coulomb-like potentials. In a semiclassical approximation to the Lagrangian of eq. (4.7), the monopole’s activity is  $(1/3!) \lambda_0 \langle \psi \rangle^3$ . As it should, the monopole’s density goes to zero when the vacuum expectation value of  $\psi$  goes to zero. This picture of confinement in the  $2 + 1$   $Z(N)$  model is similar to Polyakov’s instanton one [7].

## 6. From $2 + 1$ to $3 + 1$

This section extends the ideas in  $2 + 1$  dimensions to  $3 + 1$  dimensions, relates the  $2 + 1$  model to the  $3 + 1$  one, and indicates that proof of confinement in  $2 + 1$  is probably sufficient to prove confinement in  $3 + 1$ . This means that the  $2 + 1$  dimensional  $Z(N)$  model is more than just a toy laboratory. It is important to calculate the vortex properties and find the effective Lagrangian. Once these are known, we will probably know whether gauge theories in  $3 + 1$  dimensions confine *via* the  $Z(N)$  mechanism.

Just as instantons in  $1 + 1$  dimensions are solitons in  $2 + 1$ , the  $Z(N)$  solitons in  $2 + 1$  dimensions are strings in  $3 + 1$ . The vortex solution, which is  $x$  and  $y$  dependent in Euclidean  $1 + 1$  dimensions, becomes  $t$  independent in  $2 + 1$  and  $t$  and  $z$  independent in  $3 + 1$ . It respectively looks like a point, a line and a sheet. The latter two manifest themselves as loops and closed surfaces and their associated quanta are particles and closed strings. The ideas in Euclidean  $2 + 1$  dimensions are relevant for  $3 + 1$  dimensions because a time slice of  $3 + 1$  looks like Euclidean  $2 + 1$ . The solitons in  $2 + 1$  which were particles tracing out trajectories now become strings tracing out surfaces. Hence, *in the physical world the topological objects are closed strings manifesting themselves in Euclidean space as “ $Z(N)$  surface solitons”*.

$\star$  This is true in an “Abelized” gauge where the  $\eta$  vector’s color direction is fixed, that is, spatially constant.

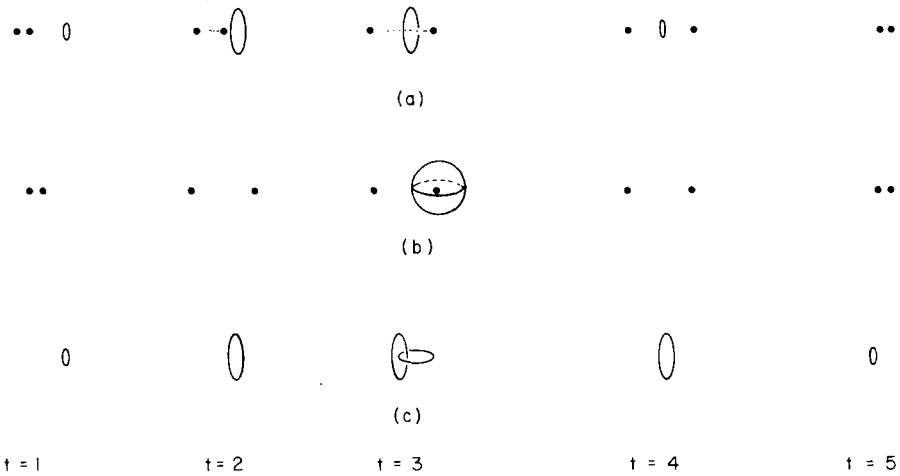


Fig. 17. Linking of a sphere and a loop in four dimensions. These three figures show a temporal sequence in which a sphere and a loop link. Each  $t = n$  value represents a time slice. In general, a time slice of a loop and a sphere yields respectively two points and a loop. Exceptions to this occur when the loop or sphere are contained within a single time slice, in which case they respectively look like a loop ((c) at  $t = 3$ ) and a sphere ((b) at  $t = 3$ ). (a) shows the generic case: a pair of “particles” and a small “closed string” are produced out of the vacuum. One of the particles shoots through the loop, which subsequently shrinks and disappears. The particles then annihilate. In (b) the sphere is contained in the  $t = 3$  slice. Again, a pair of particles is produced and the two separate. One of them is instantaneously surrounded by the sphere, which subsequently vanishes. The two then annihilate. In (c) the loop is contained in the  $t = 3$  slice. A closed string is produced. It expands; then with the sudden appearance of the loop, it links. The loop instantly disappears and the closed string shrinks and vanishes.

The key defining property is linking number. In  $2 + 1$  dimensions two non-intersecting loops can link and linking number is well-defined for oriented curves. In  $3 + 1$  dimensions a closed surface and a loop can link. Again, for oriented surfaces and oriented loops the linking number is well-defined. Fig. 17 illustrates using “time lapse photography” how a sphere  $\star$  and a loop can link. Because of this, a  $Z(N)$  topology characterizes “surface solitons”. It works just like it does in one lower dimension. An idealized, that is infinitely thin, surface soliton,  $S$ , satisfies

$$[\text{tr P exp}(ig \oint_C \mathbf{A} \cdot d\mathbf{l})] = 3 \exp(\frac{2}{3}\pi i), \tag{6.1}$$

whenever  $C$  links with  $S$ . Far from the surface where potentials are pure gauge, eq.

$\star$  Besides spheres there will be tori and other oriented closed surfaces. The most general oriented surface is topologically equivalent to a sphere with  $n$  handles (an object of genus,  $n$ ). These are allowed; the other two-dimensional surfaces, the non-orientable ones such as the projective plane, are not allowed because linking number cannot be defined.

(6.1) is valid. Near the surface of a smeared or physical surface soliton, eq. (6.1) is incorrect. We shall always use “idealized” solitons. Whenever a loop, C, links with S, there is a map from C into the gauge group (as discussed in sect. 4). These maps are characterized by  $\Pi_1(\text{SU}(3)/\text{Z}(3)) \approx \text{Z}(3)$ . Eq. (6.1) says this map is a non-trivial element of  $\Pi_1$ . Similarly, the topology can be discussed in terms of Higgs fields. Far from the surface, the Higgs fields take on values in the minimum, M, of the potential. Moving along C traces a closed curve in M. Again, these are characterized by  $\Pi_1(\text{M}) = \text{Z}(3)$ . As long as C and S are kept away from each other, there is no way to unlink them. Likewise, as one moves C, the topological element  $\Pi_1(\text{M})$  or  $\Pi_1(\text{SU}(3)/\text{Z}(3))$  cannot jump since such continuous movements are homotopies. The topology is virtually the same as in one lower dimension.

The topological surfaces will have properties similar to the 2 + 1 dimensional case.

(i) On the surface the Higgs fields become “aligned” so that at least one generator commutes with them. We expect only one generator,  $\eta^l \lambda^l$ .

(ii) For “idealized” surfaces, there is a delta-function contribution to the flux in the  $\eta$  direction. Let  $e^{(1)}(x)$  and  $e^{(2)}(x)$  be two orthonormal tangent vectors to the surface at  $x$ . Then

$$\eta^l(x) e_\alpha^{(1)}(x) e_\beta^{(2)}(x) \frac{1}{2} F_{\mu\nu}^l(x) \epsilon_{\alpha\beta\mu\nu} = \frac{2\pi}{g} \delta_\perp^2(x). \tag{6.2}$$

The vector,  $\eta^l$ , is normalized so that  $\sum_l \eta^l \eta^l = \frac{3}{4}$  and  $\delta_\perp^2(x)$  is a delta function in the variables perpendicular to the surface. More accurately,

$$F_{\mu\nu}^l(x) = \frac{2\pi}{g} \oint_{\substack{\text{over} \\ \text{surface} \\ \text{soliton}}} \frac{4}{3} \eta^l(y) \delta^4(x-y) \frac{1}{2} \epsilon_{\alpha\beta\mu\nu} dS_{\alpha\beta}. \tag{6.3}$$

(iii) The mass of the soliton in 2 + 1 dimensions is (roughly) the energy per unit length of the string. The engineering dimensions of the parameters that give the mass in 2 + 1 dimensions have dimensions of mass squared in 3 + 1; couplings acquire different dimensions in different dimensions. For example, the mass of the Nielsen-Olesen vortex due to the Higgs potential goes like  $\langle \phi \rangle^2$  [5], In 2 + 1 dimensions,  $\langle \phi \rangle^2$  has dimensions of mass, whereas in 3 + 1, it has dimensions of mass squared.

Linking number formulas in four dimensions also exist. Ref. [17] gives

$$\begin{aligned} n &= \frac{1}{2\pi^2} \oint_{x \in S} dS_{\mu\nu} \oint_{y \in C} dy_\beta \frac{1}{2} \epsilon_{\alpha\beta\mu\nu} \frac{(y-x)_\alpha}{|y-x|^4} \\ &= \frac{1}{2\pi^2} \int_0^1 d\sigma \int_0^1 d\tau \int_0^1 ds \frac{1}{2} \epsilon_{\alpha\beta\mu\nu} \frac{[y(s) - x(\sigma, \tau)]_\alpha}{|y(s) - x(\sigma, \tau)|^4} \end{aligned}$$

$$\begin{aligned} & \times \frac{dy_\beta(s)}{ds} \frac{dx_\mu(\sigma, \tau)}{d\sigma} \frac{dx_\nu(\sigma, \tau)}{d\tau} \\ & = \oint_S \frac{1}{2} B_{\mu\nu}(x) dS_{\mu\nu}(x), \end{aligned} \quad (6.4)$$

where

$$B_{\mu\nu}(x) = \frac{1}{2\pi^2} \oint_C \epsilon_{\alpha\beta\mu\nu} \frac{(y-x)_\alpha}{|y-x|^4} dy_\beta. \quad (6.5)$$

In eq. (6.4)  $S$  is a closed oriented surface,  $dS_{\mu\nu}$  is the surface element,  $C$  is a closed oriented curve,  $dy_\beta$  is its line element, and  $1/2\pi^2$  is the volume of the three-dimensional unit sphere. In the second form, the curve,  $C$ , is parametrized by  $s$  so that  $y(s)$  is the location at “time”,  $s$ ,  $y(0) = y(1)$ , and  $dy_\beta/ds$  is the “velocity”. The variables  $\sigma$  and  $\tau$  parametrize the surface in the same manner as in dual string theory. They take on values in the unit square, and  $x(\sigma, \tau)$  is the location of the surface at that  $(\sigma, \tau)$  value. Because the surface is closed,  $x(0, \lambda) = x(1, \lambda) = x(\lambda, 0) = x(\lambda, 1) =$  a constant for  $0 \leq \lambda \leq 1$ , that is, the boundary of the square is mapped to the same point. The linking number is an integer.

If one were doing static electromagnetism in  $4 + 1$  dimensions, the  $B_{\mu\nu}(x)$  of eq. (6.5) would be the magnetic field at  $x$  produced by a unit current flowing through  $C$ . It has a “vector field” gauge invariance:

$$B_{\mu\nu}(x) \rightarrow B_{\mu\nu}(x) + \partial_\mu \chi_\nu(x) - \partial_\nu \chi_\mu(x). \quad (6.6)$$

Under the transformation of eq. (6.6), eq. (6.4) is unchanged. This means that there are many other permissible forms for  $B_{\mu\nu}$ . There is also a gauge invariance for the gauge since  $\chi_\mu \rightarrow \chi_\mu + \partial_\mu \chi$  leaves eq. (6.6) invariant.

The vacuum will be a gas of  $Z(N)$  closed surfaces. This will be an interesting statistical mechanics ensemble. Such a gas might yield a field theory for strings for which the methods of sects. 2–4 could be mimicked. Even in the absence of interactions between points on the surface, where a free field theory of closed strings is expected, constructing such a field theory will be difficult\*. We are, therefore, unable to calculate the Wilson loop. Such a calculation involves summing over the gas of surfaces, weighting each surface,  $S$ , by the factor

$$\exp \left[ \frac{2}{3} \pi i \int_S \frac{1}{2} B_{\mu\nu}(y) dS_{\mu\nu}(y) \right], \quad (6.7)$$

with  $B_{\mu\nu}$  given by eq. (6.5) for  $C$  being the Wilson loop. Instead, we make the following observations and conjectures: assume that the soliton in  $2 + 1$  dimensions

\* Kaku and Kikkawa have constructed a field theory for strings. Unfortunately, it is non-covariant and hence not useful from our point of view. See ref. [30].

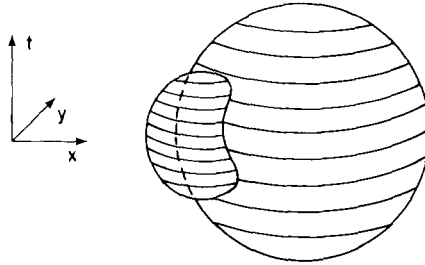


Fig. 18. A triplet surface. This surface is contained in the  $z = 0$  slice of space-time, hence the  $z$  direction is not shown. The object is like a smaller bubble stuck to a bigger one.

has a positive mass so that in  $3 + 1$  the surfaces carry a positive surface energy. The topological sector generates a theory of strings. Such closed strings do not carry quantum numbers. They occur as space-time events: a point suddenly appears, sketches into a ringlet, and shrinks away. These ringlets are restricted to be small and to last for brief durations because of the large surface action generated. An exception to this might be the tachyons in the dual string model. These two low-lying states, as minute closed strings, resemble particles of negative mass squared. They trace out long trajectories which, because there is no conserved quantum number, can begin or end in the vacuum. As long thin cylinders, they have little effect on Wilson loops except for possibly a mass renormalization; such thin tubular configurations are unlikely to link. All in all, such a system does not produce confinement.

Because of the  $Z(3)$  structure of the topology, there may exist other types of surfaces. These are the analogs of the  $2 + 1$  triplets. Fig. 18 shows a “triplet surface” imbedded in the slice,  $z = 0$ . The temporal evolution of this configuration is shown in fig. 19. Time slices sometimes yield triplets and sometimes ringlets. These string configurations are very different from interacting dual strings. The latter interact by breaking or by joining ends as well as a “four-point” interaction where two strings touch in the middle and exchange string halves. These interactions occur at a specific location at a particular time. Fig. 19 shows that triplet surfaces look like an open string circumscribed by a closed one or more accurately three open strings joined at the ends; the three are in constant interaction. Thus, this string theory is unlike anything previously considered in dual models.

Triplet surfaces contain, of course, monopoles. The intersection of the three surfaces is a curve which is to be identified with the monopole’s trajectory. Like closed soliton surfaces, triplet surfaces must also be assigned an orientation. This induces an orientation on the boundary, that is, for the monopole loop <sup>\*</sup>. Hence,

<sup>\*</sup> The definition of orientation for a manifold can be found on page 119 of ref. [31]. The way in which an oriented manifold induces an orientation for its boundary can also be found here.

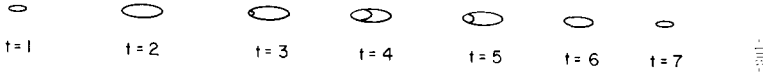


Fig. 19. The temporal sequence of fig. 18.

as should be the case, the monopole trajectories are oriented and closed, indicating a conserved quantum number (monopole charge). Since three surfaces span the monopole loop, the system at a particular time looks like a monopole-antimonopole pair joined by three vortices (fig. 8). Because these three vortices carry a positive energy per unit length, the monopole-antimonopole pair are inexorably bound by a linear potential. Triplet surfaces, comprised of the monopole loop and three spanning surfaces, must be small and hence monopole vacuum loops are rare events. These “neutral” objects have few physical effects. As in one lower dimension, there is no confinement. The interesting case of topological symmetry breakdown, where confinement is expected, will be discussed shortly. We first must show how linking number can be defined for triplet surfaces.

Eq. (6.4) no longer works for triplet surfaces. This is the higher-dimensional analog of the problem discussed in sect. 4. The resolution is similar: a singular “surface axial gauge” for  $B_{\mu\nu}$  must be used. Let  $S_C$  be any surface which spans the Wilson loop,  $C$ . Then

$$B_{\mu\nu}^{S_C}(x) = \int_{S_C} \frac{1}{2} \epsilon_{\mu\nu\alpha\beta} \delta^4(x-y) dS_{\alpha\beta}(y), \quad (6.8)$$

works. This  $B_{\mu\nu}^{S_C}$ , when substituted into eq. (6.7) and integrated over the triplet surface,  $S$ , yields the correct phase factor. The surfaces  $S_C$  and  $S$  should not be confused; the former is any surface whose boundary is the Wilson loop while the latter is the triplet surface. It is not hard to find a gauge transformation which moves  $S_C$ : let  $S_C^1$  and  $S_C^2$  be two Wilson loops spanning surfaces. They form a closed surface. Let  $V$  be any volume, i.e., a three-dimensional manifold, whose boundary is their union. For  $y \in V$ , let  $\eta_\mu(y)$  be the vector orthonormal to  $V$  at  $y$  in four-space. Then

$$\begin{aligned} \chi_\mu(x) &= \int_{y \in V} \eta_\mu(y) \delta^4(x-y) d^3y \\ &= \int_{y \in V} \frac{1}{3!} \epsilon_{\mu\alpha\beta\gamma} \delta^4(x-y) dV^{\alpha\beta\gamma}(y), \end{aligned} \quad (6.9)$$

when used in eq. (6.6) affects a gauge transformation from  $B_{\mu\nu}^{S_C^1}$  to  $B_{\mu\nu}^{S_C^2}$ . Actually, we have not defined how the sign of  $\eta_\mu$  is to be chosen. One choice gauges  $B_{\mu\nu}^{S_C^1}$  into  $B_{\mu\nu}^{S_C^2}$ ; the other gauges  $B_{\mu\nu}^{S_C^2}$  into  $B_{\mu\nu}^{S_C^1}$ . The fact that  $V$  is not unique (there are many volumes,  $V$ , whose boundary is  $S_C^1 \cup S_C^2$ ) reflects the fact that the gauge transformation is not unique (“a gauge invariance for the gauge”) as previously discussed.

What would happen if a general gauge was chosen in eq. (6.7)? Take the  $B_{\mu\nu}$  of eq. (6.8) and perform the gauge transformation  $B_{\mu\nu} \rightarrow B_{\mu\nu} + \partial_\mu \chi_\nu - \partial_\nu \chi_\mu$ . Then, each of the three surfaces comprising a triplet surface would contribute an extra factor

$$\int_S (\partial_\mu \chi_\nu - \partial_\nu \chi_\mu) dS_{\mu\nu} = \int_{\partial S} \chi_\mu dx^\mu = \oint_{C_M} \chi_\mu dx^\mu . \tag{6.10}$$

The boundary of each surface is  $C_M$ , the monopole loop. Triplet surfaces would get multiplied by the unwanted phases

$$\exp(3i \oint_{C_M} \chi_\mu dx^\mu) . \tag{6.11}$$

This is innocuous if  $\oint_{C_M} \chi_\mu dx^\mu = \frac{2}{3}\pi n$ . For a gas of triplet surfaces, we must require that

$$\oint_C \chi_\mu dx^\mu = \frac{2}{3}\pi n , \tag{6.12}$$

for all loops, C. By shrinking C to a point in a plane, P, eq. (6.12) implies that  $\partial_\mu \chi_\nu - \partial_\nu \chi_\mu$  must be singular on the plane,  $P^\perp$ , perpendicular to P. Hence gauge transformations can only move  $S^C$  around and eq. (6.8) is the most general form for  $B_{\mu\nu}$ ; one cannot smooth  $B_{\mu\nu}$  out.

The interesting case is when topological symmetry breakdown occurs, that is, when the vortices in 2 + 1 dimensions have a negative (or perhaps zero) mass squared. Then, topological surfaces are expected to have a negative surface action density and will populate the four-dimensional world. This implies that the closed topological strings will have a negative Regge slope, a thought that, at first, seems preposterous because of the infinite number of tachyons. However, the situation is not as bad as it appears and even has a simple physical interpretation. First of all, a total collapse is not expected. The same repulsive vortex forces which stabilize the proliferation of closed loops in 2 + 1 dimensions will be present in 3 + 1 dimensions

Table 1

Wilson lattice gauge theory	SU(3) gauge theory
electric fields	magnetic fields
quarks	monopoles
closed electric string ↔ electric fluctuations	closed magnetic string ↔ magnetic fluctuations
positive slope ↔ zero or negative slope	positive slope ↔ zero or negative slope



and will stabilize the vacuum <sup>\*</sup>. Secondly, triplet surfaces will become monopoles with their magnetic flux spread out. Consider, for example, a large monopole loop with its three topological  $Z(3)$  surfaces. Unlike the case,  $m^2 > 0$ , where the surfaces are the ones of minimal surface area, the surfaces can be anywhere. The quantum mechanical sum over all possibilities will make the flux evenly distributed rather than being focused in tubes. We conjecture that a negative slope parameter for these types of strings results in “fields”. By this, we mean a string theory gets transformed into a field theory. There is another example of this phenomenon which should clarify what we mean, namely Wilson’s lattice gauge theory [33]. Keep the lattice spacing finite. In the strong coupling limit the electric field is focused into tubes. States of two quarks connected by an electric flux tube or a torus of flux are permissible. The theory has strings with positive slope parameters. As the coupling constant is lowered, the electric flux begins to spread out more and more until a phase transition occurs where it spreads out uniformly in the usual Coulomb-like manner. At the phase transition, spontaneous symmetry breakdown of electric strings has occurred. The effective surface action density (including entropy contributions) has become zero. We conjecture that the phase transition from  $\langle \psi \rangle = 0$  to  $\langle \psi \rangle \neq 0$  is the same phenomenon except with “dual” magnetic fields (see table 1). This is what we mean by strings of negative Regge slope being metamorphosed into fields. When this happens topological surfaces become magnetic fluctuations and monopoles bound in dipoles become liberated. These monopoles will now be instrumental in confining quarks. *If  $m^2 < 0$  for 't Hooft vortices and triplets exist in  $2 + 1$  dimensions, then in  $3 + 1$  dimensions monopoles, previously confined in monopole-antimonopole pairs by magnetic flux tubes, get liberated resulting in a monopole plasma.*

We conjecture that the phenomena exhibited in  $2 + 1$  dimensions will be present in  $3 + 1$ . Already discussed are magnetic fluctuations due to close surface solitons and monopoles due to surface triplets. Also guaranteed is that the non-perturbative potential will be “periodic in charge” because of eqs. (6.4) and (6.7). This screening phenomenon means that integral hypercharge multiplets will not be confined. Any approximation scheme to a Wilson loop calculation should be able to reproduce this phenomenon. Next, we expect that topological symmetry breakdown will restore the symmetry associated with H. The reason is the same as in one lower dimension: the  $Z(N)$  surfaces will fill up the vacuum until overlap repulsive forces take over. Since H is restored on these surfaces, H will be restored virtually everywhere. Finally, several arguments show why confinement in  $3 + 1$  dimensions will occur.

(i) The vacuum is a gas of monopoles. Roughly speaking, such a system confines because of a “dual Meissner effect” [8,34]. Just as a gas of current loops confines monopoles in a superconductor, a gas of magnetic current loops confines charges.

<sup>\*</sup> Bardakci has explicitly shown this in a string model [32].

(ii) The Wilson loop calculation involves a system with an area constraint (the “surface axial gauge” constraint of eq. (6.8)). If a non-trivial situation exists, the action must go as the area.

(iii) A time slice of 3 + 1 dimensions looks like 2 + 1 dimensional Euclidean space. Confinement in 2 + 1 dimensions might indicate confinement in 3 + 1. Putting a quark loop in a time slice appears to reduce the calculation to one lower dimension. Let us define the term, *dimensional reduction*, as when a Wilson loop in such a time slice of 3 + 1 can be calculated in 2 + 1 as a static situation. Dimensional reduction does not always occur. Consider, for example, the lower-dimensional analog: a comparison of the 1 + 1 dimensional instanton gas of Nielsen-Olesen vortices to the Nielsen-Olesen model in 2 + 1 dimensions with vortex topological symmetry breakdown. The former looks approximately like a time slice of the latter. However, sect. 3 obtained a logarithmic potential for the 2 + 1 model, whereas a linear potential occurs in 1 + 1 dimensions. The reason for this is clear. When the Wilson loop is placed in a time slice, the  $B_\mu$  of eq. (3.6) is non-zero for all times. The Wilson loop affects the vortex gas throughout the 2 + 1 dimensional world and not just at one time. Dimensional reduction does not happen. If, however,  $B_\mu$  were zero outside the time slice, it would happen. This is the case for the 't Hooft model with triplets. The  $B_\mu$  of eq. (4.12) has its support in the time slice containing the Wilson loop. This is why linear confinement occurs: the calculation dimensionally reduces to the 1 + 1 instanton calculation, which is known to confine. Returning to the physical world with triplet surfaces absent, the  $B_{\mu\nu}$  of eq. (6.5) is non-zero for all times. For this situation, the calculation does not dimensionally reduce. If  $B_{\mu\nu}$  were forced to be in a gauge with the support of  $B_{\mu\nu}$  in a time slice, then topological symmetry breakdown would yield the logarithmic potential obtained in 2 + 1 dimensions, but this is not the case. Our guess is that the action will go like  $\int B_{\mu\nu}^2 d^4x$  so that closed surfaces yield only a  $1/r$  potential. When there are triplet surfaces the situation is completely different;  $B_{\mu\nu}$  is forced to be in a “surface axial gauge” such as eq. (6.8), the calculation dimensionally reduces and a linearly confining potential (the same one as in sect. 4) is obtained. With triplet surfaces present, a time slice of the real world does indeed look like the 2 + 1  $Z(N)$  model with triplets and confinement.

### 7. Relation to Mandelstam’s scheme

There is a similarity between Mandelstam’s confinement scheme [8] and topological symmetry breakdown with triplet surfaces. We shall touch upon the common points and differences. Here is a quick review of his quark confinement.

(a) The Coulomb gauge contains a term  $E_L^\alpha \cdot E_L^\alpha$  in the Hamiltonian where

$$E_L^\alpha = \nabla[\nabla^2 \delta^{\alpha\gamma} - gf^{\alpha\beta\gamma} A_i^\beta \nabla_i]^{-1} \rho^\gamma . \tag{7.1}$$

The inverse operator,  $[\nabla^2 \delta^{\alpha\gamma} - gf^{\alpha\beta\gamma} A_i^\beta \nabla_i]^{-1}$ , akin to  $|ax - b|^{-1}$  in a one-dimensional quantum system, will produce infinities unless the vacuum is suitably chosen.

(b) In the  $A_z = 0$  axial gauge, this problem becomes equivalent to whether local  $(x, y)$  dependent (but  $z$ -independent) gauge transformations annihilate the vacuum. The bare vacuum fails to do this and makes a poor starting point to perturb around.

(c) A vacuum comprised of monopoles has the right properties: the Dirac tubes (in the  $A_z = 0$  gauge, the Dirac strings become tubes for finite-sized smeared-out monopoles) produce random gauge rotations. Such a state will be a singlet under local  $(x, y)$  dependent gauge transformations. It is a candidate for the vacuum state.

(d) Such a monopole gas confines in not so different a way from the Polyakov model [7].

In short, Mandelstam's gas results in (i) a restoration of  $SU(3)$  gauge symmetry and (ii) quark confinement.

Likewise, we have shown similar results when topological symmetry breakdown occurs in the presence of triplets: a monopole vacuum is generated and at least a global  $U(1)$  color symmetry is restored. We have been unable to prove the restoration of the complete  $SU(3)$  gauge invariance. Like Mandelstam's vacuum, confinement is a consequence. The ideas of sects. 4–6 neatly jell with Mandelstam's.

We have given support to Mandelstam's work in showing how monopoles naturally arise in an  $SU(3)$  gauge theory. The type of monopoles Mandelstam has been using (Wu-Yang ones [35]) are probably different from those in an  $SU(2)$  gauge theory. For  $SU(3)$  our monopoles carry  $6\pi/g$  units of flux. We have also elucidated on the dynamics of the system; in particular, how negative Regge slopes transform magnetic flux tubes into magnetic fields. We have not shown why topological symmetry breakdown takes place. If Mandelstam is correct then he has given us the reason: such a breakdown occurs because of the  $|\alpha x - b|^{-1}$  problem. It is a matter of symmetry, monopoles and disorder. In essence, the difference between no confinement and confinement is the difference between an ordered system with a broken symmetry and a disordered system with monopoles acting as the symmetry restoring agent.

## 8. Open questions

(A) How does the phase diagram look for an 't Hooft  $SU(N)$  gauge theory? In particular, how many phases are there? In fig. 4, we have drawn three phases, corresponding to (a) a normal boson potential, (b) a Higgs boson potential, and (c) a Higgs soliton potential. 't Hooft conjectured that phases (a) and (c) are the same. He argues that the soliton's mass squared must be proportional to the Higgs mass squared, the proportionality sign being negative since when the Higgs bosons are tachyonic the solitons have a positive mass. Analyticity implies that the solitons are tachyons when the bosons have a positive mass squared. We feel this argument needs further justification since phase transitions induce non-analyticities. This remains an open question. In general, it is important to determine the phase dia-

gram so that the confinement phase (if it exists) and the coupling constant values which yield this phase can be found <sup>\*</sup>.

(B) Can a string field theory be constructed to describe the statistical ensemble of topological surfaces? As previously discussed, with such a construction the calculations of sects. 2–4 could be performed in 3 + 1 dimensions.

(C) What effects do instantons have? Callan, Dashen and Gross have shown that instantons create a paramagnetic vacuum which tends to expel electric fields [2]. This will surely affect the dynamics of topological surfaces. Instantons might aid in (or even cause) the topological symmetry breaking. Roughly speaking, the surfaces should couple to instantons because monopoles and associated magnetic fields are long ranged. They can contribute to

$$\int F_{\mu\nu} \tilde{F}^{\mu\nu} d^4x, \quad (8.1)$$

in contrast to short-ranged field configurations which cannot be due to the fact that eq. (8.1) can be written as a surface integral.

(D) Are Higgs bosons necessary? As indicated in sects. 3 and 4, topological vortices can be characterized solely in terms of gauge potentials by using path-ordered products. This suggests that, perhaps, they exist independently of the Higgs fields. We feel this may be the case for an  $SU(N)$  gauge theory but certainly not for a  $U(1)$  theory. The Higgs bosons in the 2 + 1 dimensional Nielsen-Olesen model serve two important purposes:

- (i) they smooth out the short-distance singularities which would otherwise occur;
- (ii) they ensure that  $\exp(i e \int_x^y A \cdot dl)$  returns to 1 when  $y$  loops around the vortex and returns to  $x$ .

Purpose (i) is not as important as purpose (ii) for the existence of vortices. Purpose (i) is a short-distance phenomenon which might be cured quantum mechanically or through renormalization. In the absence of Higgs fields, there are still finite-energy vortex configurations (obtained by using the classical values of  $A_\mu(x)$  when the Higgs fields were present) although they are not energy minima. In contrast purpose (ii) is essential. With Higgs bosons, the Higgs field must take on values in the minima of the potential and be covariantly constant far from the vortex. This means that  $\exp(i e \int_x^y A \cdot dl)$  is precisely the rotation (in this case, phase) which transforms the classical Higgs field at  $x$  to its value at  $y$ . As one goes around the loop and returns to  $x$ , this guarantees that  $\oint A \cdot dl = 2\pi n/e$ . This means that magnetic flux gets quantized. Without Higgs fields, configurations such as  $A_\phi(x) = a(\rho)/\rho$ , with  $a(\rho) = 0$  at the vortex and  $a(\rho) = c/e$  far away, have finite energy. Here,  $c$  is an arbitrary constant. This means that flux is not quantized and  $\int A \cdot dl = 2\pi c/e$  is arbitrary. The solitons lose their identity as tubes of conserved quantized flux. Instability occurs because tubes of  $2\pi/e$  units can dissolve into many smaller tubes,

<sup>\*</sup> Ref. [36] is of interest in this connection.

say  $n$  tubes of  $2\pi/en$  units. Thus purpose (ii) is the essential stabilizing effect of Higgs bosons in the Nielsen-Olesen  $U(1)$  model.

In the non-Abelian case, purpose (i) still functions but it is possible that property (ii), which is now modified to

$$[\text{P exp}(ig \oint_C \mathbf{A} \cdot d\mathbf{l})]_{\alpha\beta} = [\text{a center of the group element}]_{\alpha\beta} , \quad (8.2)$$

for loops,  $C$ , which encircle the vortex, holds even in the absence of Higgs fields. This is because Yang-Mills theories are self-interacting. These self-interactions possibly act as a replacement for the Higgs boson gauge field interactions. If a flux tube, at some instance, points in the  $z$  direction, thus contributing to  $\eta^I B_z^I$ , two of the three gauge potentials might act like the two Higgs fields while the third is the flux-generating gauge potential. This point needs further clarification, but Yang-Mills theories offer the aesthetically pleasing possibility of eliminating Higgs fields without ruining any of the physical results discussed in this paper.

If the Higgs fields are eliminated, the problem of defining a monopole returns. Previously, the vector,  $\eta^I$ , constructed out of Higgs fields, was used. We now must manufacture an  $\eta^I$  using gauge potentials. There are many ways of doing this: at each point,  $x$ , attach a closed loop,  $C_x$ . Define

$$[\text{exp}(\eta^I(x) \frac{1}{2}\lambda^I)]_{\alpha\beta} = [\text{P exp}(\oint_{C_x} \mathbf{A} \cdot d\mathbf{l})]_{\alpha\beta} . \quad (8.3)$$

In eq. (8.3) the path ordering starts at  $x$ , proceeds along  $C_x$ , and ends at  $x$ . Beginning at any other starting point is not possible. The vector,  $\eta^I(x)$ , transforms in the octet representation under gauge transformations. Using the magnetic field,  $B_i(x) = \frac{1}{2}\epsilon_{ijk}\eta^I(x)F_{jk}^I(x)$ , one can “test” for monopoles by integrating  $\int B \cdot dS$  over closed surfaces. Of course, this method generates an infinite number of magnetic fields, all of which are gauge invariant. Only a prudent choice of the  $C_x$  will yield a  $B_i(x)$  with the desired properties, that is, that this magnetic field be the one contained in vortex flux tubes and be the one associated with the monopoles contained in triplet surfaces.

(E) Is there a non-Abelian dual group? Implicit in our discussion is the dualities between magnetic fields and electric fields, monopoles and charges, etc. Table 1 illustrates some of these. When there are Higgs bosons and a unique  $\eta^I$ , i.e.,  $H = U(1)$ , the symmetry group of the surfaces (and monopoles) is either  $U(1)$  or  $Z(3)$  depending on whether triplets exist. If the Higgs fields can be eliminated, will there be a non-Abelian dual group? This question will be hard to answer because of the quantum mechanical effects inherently associated with an internal symmetry group, namely, color zitterbewegung [22]. This goes beyond our discussions which have been classical in character.

## 9. Summary

To prove confinement the following must be shown:

(a) that the 't Hooft  $Z(N)$  solitons in  $2 + 1$  dimensions have a phase with  $m^2 < 0$  and that there are repulsive forces which stabilize the vacuum;

(b) that  $\lambda_0(\psi^3 + \psi^{*3})$  (or similar) terms exist, that is, triplets are present;

(c) that a time slice of the four-dimensional world is described by the physics of the  $2 + 1$  dimensional model.

When the above are satisfied, we expect:

(i) confinement;

(ii) restoration of some or all of the original color symmetries.

Without (b) topological symmetry breakdown gives only a  $\ln r$  potential in  $2 + 1$  dimensions and (most likely) a  $1/r$  potential in  $3 + 1$ . Showing (a) and (b) is the next calculational step. These problems can be approached by using:

(i) Semiclassical methods [37]. These can determine how the soliton's mass depends on the parameters in the Higgs potential.

(ii) A field theory for solitons [13,21]. This should be useful in determining the forces between vortices and whether the vacuum stabilizes for  $m^2 < 0$ .

(iii) Mandelstam's operator methods [8]. These should be applied to  $Z(N)$  type monopoles. These methods might be useful in determining vacuum instabilities and hence why the soliton's mass is negative.

(iv) Halpern's dual field strength formulation [38]. This approach might exhibit the topological solitons and their properties directly. It might also be helpful in determining whether vacuum instabilities exist.

Most difficult will be showing (b), that triplet configurations exist, although there is no *a priori* reason (in the sense of a conservation law) why they shouldn't.

Additional problems particular to four dimensions are:

(a) Whether the closed soliton strings have negative Regge slopes and whether this makes sense as this paper suggests.

(b) Whether triplet surfaces occur. A new type of dual string model is needed.

(c) Whether the Higgs potential can be done away with. Are there singular but stable topological solitons in pure Yang-Mills theories?

(d) If (c) holds, there are no mass scales. How does dimensional transmutation [39] \* come about?

We wish to emphasize the following points:

(a) The non-Abelian (and to a lesser extent non-perturbative) nature of the confinement. The selfsame method can at most yield a  $\ln r$  potential for a  $U(1)$  theory. This is because monopoles, i.e., triplets, cannot occur in an Abelian theory and they are essential in the linear confinement.

\* Thorn has presented a pedagogical example of dimensional transmutation in the two-dimensional quantum mechanics problem with a potential  $g\delta^2(x)$ . See ref. [40].

(b) The potential between quarks is expected to have (i) a linear piece due to triplet surfaces, (ii) most likely a  $1/r$  non-perturbative large- $r$  piece due to closed surfaces, and (iii) a (possibly screened or antiscreened) perturbative piece which dominates the short-distance physics.

Let us conclude by summarizing what we have shown.

(i) In  $2 + 1$  dimensions we have exhibited how to do calculations when topological symmetry breakdown occurs by using a macromolecule analogy. We believe this will form the prototype of future topological symmetry breakdown calculations.

(ii) We have extended 't Hooft's confinement scheme in  $2 + 1$  dimensions to  $3 + 1$ . Previously, only an operator algebra yielding different phases was obtained.

(iii) We have discussed the dynamics of quark confinement, namely, how monopoles naturally arise in a non-Abelian gauge theory; how negative Regge slopes make sense, change  $Z(N)$  strings into magnetic field fluctuations, generate the usual  $1/r$  radial monopole magnetic field and liberate the monopoles (of the bound monopole-antimonopole pairs) which are so instrumental to quark confinement.

(iv) We have connected the physics of the  $Z(N)$  models to Mandelstam's confinement scheme.

*In short, we believe we are at the verge of proving confinement in non-Abelian gauge theories.*

The author thanks Korkut Bardakci for delightful discussions, particularly on the properties of the  $Z(N)$  vortex. Korkut Bardakci's polemical questions often forced the author to clarify to himself his ideas and, in one instance, this resulted in a revamping of the author's thinking. The author thanks him for his excellent guidance. I would like to thank Joe Polchinski for discussions concerning the nature of The Saddle point solutions.

## References

- [1] G. 't Hooft, Nucl. Phys. B138 (1978) 1.
- [2] C.G. Callan, R.F. Dashen and D.J. Gross, A theory of hadronic structure, Princeton preprint (June, 1978).
- [3] E. Witten, Nucl. Phys. B149 (1979) 285.
- [4] T. Yoneya, Nucl. Phys. B144 (1978) 195;  
A. Casher, Nucl. Phys. B151 (1979) 353.
- [5] H.B. Nielsen and P. Olesen, Nucl. Phys. B61 (1973) 45.
- [6] T. Banks, R. Myerson and J. Kogut, Nucl. Phys. B129 (1977) 493.
- [7] A.M. Polyakov, Nucl. Phys. B121 (1977) 429.
- [8] S. Mandelstam, Confinement in non-Abelian gauge theories, talk given at the Washington Meeting of the American Physical Society (April, 1977).

- [9] M. Stone and P.R. Thomas, Phys. Rev. Lett. 41 (1978) 351.
- [10] S. Samuel, Phys. Rev. D18 (1978) 1916.
- [11] K.F. Freed, Adv. Chem. Phys. 22 (1972) 1;  
F.W. Wiegel, Phys. Reports 16 (1975) 59; and references therein.
- [12] M.B. Halpern and P. Senjanovic, Phys. Rev. D15 (1977) 1655;  
M.B. Halpern and W. Siegel, Phys. Rev. D16 (1977) 2486;  
M.B. Halpern, A. Jevicki and P. Senjanovic, Phys. Rev. D16 (1977) 2476.
- [13] K. Bardakci and S. Samuel, Phys. Rev. D18 (1978) 2849.
- [14] R.J. Rubin, J. Chem. Phys. 20 (1952) 1940; J. Math. Phys. 8 (1967) 576;  
S.F. Edwards, Proc. Phys. Soc. 85 (1965) 613; 88 (1966) 265;  
C. Domb, Adv. Chem. Phys. 15 (1969) 229.
- [15] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Yu. S. Tyupkin, Phys. Lett. 59B (1975) 85.
- [16] P. Goddard and D.I. Olive, New developments in the theory of magnetic monopoles, CERN preprint TH.2445 (1978).
- [17] H. Flanders, Differential forms (Academic Press, New York, 1963) p. 74.
- [18] E.M. Purcell, Electricity and magnetism (McGraw-Hill, New York, 1965) ch. 6, sect. 1.
- [19] L.I. Schiff, Quantum mechanics, third edition (McGraw-Hill, New York, 1968), sect. 27, p. 196.
- [20] C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (Freeman, New York, 1973) p. 1149.
- [21] K. Bardakci, Field theory for solitons, II, Berkeley preprint (Dec., 1978).
- [22] S. Samuel, Nucl. Phys. B149 (1979) 517.
- [23] P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. B125 (1977) 1.
- [24] J.S. Langer, Ann. of Phys. 41 (1967) 108; 54 (1969) 258; Proc. IUPAP Conf. on Statistical Mechanics, Amsterdam, 1973;  
F.W. Wiegel, ref. [11].
- [25] S. Coleman, Phys. Rev. D15 (1977) 2929.
- [26] J.M. Kosterlitz and D.J. Thouless, J. Phys. C6 (1973) 1181.
- [27] A.M. Polyakov, JETP Lett. 20 (1974) 194;  
G. 't Hooft, Nucl. Phys. B79 (1974) 276;  
P. Goddard and D.I. Olive, ref. [16].
- [28] S. Coleman, in Classical lumps and their quantum descendants, Proc. 1975 Int. School of Physics, "Ettore Majorana".
- [29] N.H. Christ, Phys. Rev. Lett. 34 (1975) 355.
- [30] M. Kaku and K. Kikkawa, Phys. Rev. D10 (1974) 1110.
- [31] M. Spivak, Calculus on manifolds (Benjamin, New York, 1965) p. 119.
- [32] K. Bardakci, private communication.
- [33] K.G. Wilson, Phys. Rev. D10 (1974) 2445.
- [34] S. Mandelstam, Phys. Reports 23 (1976) 245;  
G. 't Hooft, Proc. EPS Int. Conf. on High-energy physics, Palermo, June 1975, ed. A. Zichichi (Editrice Compositori, Bologna, 1976) p. 1225.
- [35] T.T. Wu and C.N. Yang, in Properties of matter under unusual conditions, ed. H. Mark and S. Fernach (Interscience, New York, 1969) p. 349.
- [36] T. Banks and E. Rabinovici, Finite temperature Abelian lattice gauge theories, IAS preprint (Jan., 1979).
- [37] J. Goldstone and R. Jackiw, Phys. Rev. D11 (1975) 1486;  
R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D10 (1974) 4115, 4130, 4138; 11 (1975) 3424;  
R. Jackiw, Rev. Mod. Phys. 49 (1977) 681; R. Rajaraman, Phys. Reports 21 (1975) 227.



- [38] M.B. Halpern, Field-strength and dual variable formulations of gauge theory, LBL preprint 7980 (Sept. 1978).
- [39] S. Coleman and E. Weinberg, *Phys. Rev. D* 7 (1973) 1888.
- [40] C. Thorn, Talk given at Lawrence Berkeley Laboratory; see also Quark confinement in the infinite momentum frame, MIT preprint (Sept., 1978).