PHYSICAL REVIEW D VOLUME 19, NUMBER 8 15 APRIL 1979

# Charge-monopole duality and the phases of non-Abelian gauge theories

S. Mandelstam
*Department of Physics, University of California, Berkeley, California 94720*
(Received 8 December 1978)

It is shown that the O(3) non-Abelian gauge theory possesses charge-monopole duality in the sense that, analogous to the ordinary magnetic vector potentials, one can construct O(3) "electric vector potentials," gauge-invariant combinations of which constitute the physical variables. However, while the Hamiltonian could in principle be expressed in terms of such variables, it would be very complicated; the definition of the potentials is thus not unique. The O(3) magnetic gauge group is separate from the ordinary electric gauge group. One can also formulate the theory in terms of electric (Wilson) or magnetic (Nielsen-Olesen) flux loops; the property of the loops which identifies the electric or magnetic gauge group as O(3) is studied. The possible phases of non-Abelian gauge theories are discussed briefly. The phase with real massless gluons, if it exists at all, is more complicated than the corresponding phase in an Abelian theory in the sense that it is not known how to construct a trial vacuum state without an infrared energy divergence. This phase should not be distinguished from the other phases by the absence of symmetry breaking, again in contradistinction to the Abelian gauge theory. The phase with complete Higgs symmetry breaking and that with confinement are electric-magnetic duals of one another. The relation between the Wilson condition, applied to loops at a fixed time, and the confinement of infinitely massive quarks is studied; it is hoped that the analysis will be helpful in constructing wave functions for hadronic states with confined quarks. It is suggested that Weinberg-Salam-type models may confine due to instanton effects though, for $\alpha$ around 1/137, this has no practical significance.

## I. INTRODUCTION

Non-Abelian gauge theories appear to treat charges and monopoles on a very different footing. The symmetry underlying the theory is a charge symmetry, but there seems to be no analogous symmetry associated with monopoles. Nevertheless, evidence for an underlying symmetry between charges and monopoles has gradually been increasing. Wu and Yang[1] found that one could obtain a potential distribution with a magnetic field varying like $1/r^2$ at large distances. 't Hooft[2] and Polyakov[3] showed that one could obtain a monopolelike solution of the static field equations in the presence of a Higgs field. The asymptotic behavior of their potentials was identical to that of Wu and Yang. With a suitable gauge choice, the magnetic field at infinity had the precise monopole form

$$F_{ij}^{\alpha}=\delta^{\alpha 3}\epsilon_{ijk}\frac{r^k}{|r|^3}\,. \tag{1.1}$$

In an SU($n$) theory the strength of the monopole was $n$ Dirac units.

Nielsen and Olesen[4] proposed that the "strings" of the dual model might be representations or idealizations of vortices of magnetic flux in a superconducting vacuum. Such vortices could only exist if the gauge symmetry was completely broken by Higgs fields, and they would have to contain fewer than $n$ Dirac units of flux. The model could allow open strings only if Dirac monopoles were present to absorb the flux at the ends. Nambu[5] emphasized that monopoles could only exist in the superconducting vacuum if they were joined by quantized vortices of magnetic flux, and that, in the Abelian model which he considered, they would be confined in monopole-antimonopole pairs. Mandelstam[6] extended his consideration to non-Abelian models.

't Hooft and Kogut and Susskind[7] then suggested that color confinement would occur if *electric* flux were squeezed into vortices. 't Hooft proposed a nonrenormalizable model in which such a phenomenon might occur.

The question arises whether the electric flux confinement proposed in Ref. 7 could occur by a mechanism similar to the magnetic flux confinement of Ref. 4. Mandelstam[8] suggested that one might construct a "magnetic Higgs vacuum" as a coherent plasma of Wu-Yang monopoles, just as an ordinary electric Higgs vacuum was a coherent plasma of charges. Such a vacuum would be expected to confine electric flux in the same way as an electric Higgs vacuum confines magnetic flux. 't Hooft[9] independently made a related suggestion. He considered a gauge theory with partial Higgs symmetry breaking, and a remaining unbroken Abelian gauge group. If the 't Hooft-Polyakov monopoles in such a phase became tachyonic, a new phase with confinement properties might result.

't Hooft[10] later suggested that the properties of the Higgs and confinement phases might be expressed in terms of the operators which created vortices of electric and magnetic flux. The former are the Wilson-loop operators, the latter the operators which create Nielsen-Olesen vortices. He showed that, in a phase without massless particles, the vacuum-expectation value of at least one of these operators should decrease like the exponential of the area of the loop as the loop became large. The phase in which the Nielsen-Olesen operator decreased like $e^{-A}$ was the Higgs phase, that in which the Wilson operator decreased like $e^{-A}$ was the confinement phase.

In view of all these points of analogy between electric and magnetic quantities, one is tempted to enquire whether a non-Abelian gauge theory possesses a complete electric-magnetic (or charge-monopole) symmetry. The answer to this question would give us more insight into the phases of the theory and, in particular, into the confinement phase.

If we ask the question from a dynamical point of view, i.e., if we ask whether we can write a reasonably simple Lagrangian in terms of "electric vector potentials"[11] instead of the usual magnetic vector potentials, the answer appears to be negative. In this respect non-Abelian theories differ from Abelian theories. The *elementary* quarks and gluons in non-Abelian theories definitely have electric rather than magnetic color degrees of freedom.

However, we may also distinguish between a gauge theory and any other theory kinematically, i.e., without reference to a specific Lagrangian. A gauge theory cannot be defined in terms of local, physical variables. One can introduce local potentials $A_\mu^\alpha$, but these should be regarded as auxiliary quantities. Only gauge-invariant combinations thereof constitute the physical operators. Alternatively, one can formulate the theory entirely in terms of physical, nonlocal operators, such as fixed-time Wilson-loop operators or, for Abelian theories, operators which create charged particles with their Coulomb field.

We should expect the kinematic definition of a guage theory to have direct experimental consequences. In a phase with massless particles one can easily detect such particles and the associated long-range interaction between charges and monopoles. In a phase without massless particles one can detect the linearly rising Regge trajectories associated with high-angular-momentum electric or magnetic flux loops; the centrifugal force would oppose the collapse of the loops. In the presence of quarks the leading trajectories would probably be associated with open flux tubes stretched between confined quarks.

We wish to show that, from the kinematical point of view, gauge theories are completely symmetric in electric and magnetic quantities. We can define "electric vector potentials,"[11] analogous to the ordinary magnetic vector potentials, or we could define a set of Nielsen-Olesen loop variables, analogous to the Wilson loop operator. We could construct all operators in the Hilbert space in terms of these operators; if we used the electric vector potentials, only gauge-invariant combinations thereof would constitute physical operators. The Hamiltonian, expressed in terms of our new variables, would be a very complicated function. The definition of our new variables is by no means unique; in any quantum theory one only obtains a more or less unique set of fundamental variables if one demands that the Hamiltonian adopt a simple form when expressed in terms of them.

We adopt, as our starting point, the creation operators for Nielsen-Olesen loops proposed by 't Hooft. In order to use such vortices for the construction of a complete set of "magnetic" operators we must answer several questions, such as the following:

(i) How can we construct a gauge-invariant operator which creates a Nielsen-Olesen loop in a non-Abelian gauge theory?

(ii) What property of Wilson-loop or Nielsen-Olesen-loop operators distinguishes between different gauge groups? We shall ask for the property which identifies the gauge group as O(3), since we shall not consider more general groups.

Before we can examine the questions just posed, we have to discuss some general properties of gauge theories, which we shall do in the following section. It is a nontrivial, and thus far unsolved, problem to construct a trial vacuum state with finite energy in a non-Abelian gauge theory.[12] We refer to states within the physical Hilbert space rather than to states in the enlarged Hilbert space associated with unphysical gauges. The infinities we have in mind are not ultraviolet infinities; they remain if we apply an ultraviolet cutoff and cannot be removed by renormalization. We have given arguments to indicate that the problem might be solved by constructing a magnetic Higgs vacuum mentioned above,[13] and we hope to elaborate on the subject in a future paper. For our present purpose, the important aspects of this question pertain to the residual gauge group associated with physical gauges, such as the axial gauge which we shall use in this paper.

We shall show that the correct vacuum, or any trial vacuum with finite energy density, must be invariant under transformations in the residual gauge group, a result first indicated by Schwinger.[14] As a consequence, we shall show that any operator with finite matrix elements between physical states must be invariant under such transformations. The latter property is also true if we use the temporal gauge (which is not a physical gauge); operators with finite matrix elements between physical states must then be invariant under all fixed-time gauge transformations.

In Sec. III we shall discuss the kinematical definition of a gauge theory. We shall find the property of the Wilson-loop operators which distinguishes an O(3) gauge theory from any other theory. The possible color vibrational modes of electric flux loops in a confined theory are associated with the property in question.

In Sec. IV we discuss the construction of operators which create closed Nielsen-Olesen vortices. Since the operators must be gauge invariant, the color direction of the magnetic flux will be defined with respect to an internal coordinate, i.e., with respect to some field variable, which need not be an elementary field. The simplest Nielsen-Olesen operators possess the commutation relations demanded by 't Hooft,[10] but the totality of operators of this type is not sufficient to reconstruct all operators in the theory. They are the magnetic analogs of electric loop operators constructed from the 't Hooft[2] tensor, rather than of the Wilson loops. One can then proceed to construct magnetic loop operators which are analogous to the Wilson loops; they require two internal coordinates to specify directions in color space. Rotations of such coordinates constitute an O(3) magnetic gauge group which is completely independent of the electric gauge group.

From the magnetic vortex operators we could in principle construct a set of "electric vector potentials," in terms of which all other operators could be constructed. We should note that the electric and magnetic variables are not related by an equation of the form $\tilde{F}^{\alpha,\mu\nu}=\epsilon^{\mu\nu\rho\sigma}F_{\alpha,\rho\sigma}$, since the color indices do not correspond. The term "electric vector potentials" is justified in the sense that a loop constructed from them in the manner of Wilson obeys the 't Hooft commutation relations with the actual Wilson loops. The O(3) character of the magnetic gauge group implies that the color vibrational modes of Nielsen-Olesen loops, in a phase where such loops are stable, are exactly the same as those of Wilson loops.

We should emphasize that we make no claim for the practical utility of our magnetic operators. We introduce them to exhibit the kinematical electric-magnetic symmetry of gauge theories. For actual calculations we require the elementary-field operators, which in this case are the electric operators.

Our construction of the magnetic operators verifies, and makes more precise, the conjecture of Goddard, Nuyts, and Olive[15] regarding the magnetic gauge group in the special case of O(3). We should, however, add the cautionary remark that neither electric nor magnetic gauge invariance is a symmetry in the usual sense of the word; even global gauge invariance cannot be treated as a symmetry except in Abelian theories.

In the final section we apply our analysis to make some remarks on the phases of non-Abelian gauge theories (without "sea" quarks). We also discuss the relation between the Wilson criterion and the inability of flux to spread out. Our treatment differs from that of Wilson himself in that we examine the energy of states at a given instant of time, and we apply the criterion to fixed-time loops. It is hoped that our arguments will indicate how to apply the Wilson criterion to construct trial wave functions for hadronic states with confined quarks.

## II. PHYSICAL GAUGES AND GAUGE-THEORY VACUUMS

In this section we shall demonstrate the following properties of non-Abelian gauge theories.

(i) It is a nontrivial problem to find a vacuum with finite energy density (even with an ultraviolet cutoff).

(ii) We cannot really construct physical, translation-invariant gauges at all. Such gauges have a residual gauge invariance, and matrix elements of quantities which depend on the residual gauge choice will not be finite. Particular gauges are very useful for the purpose of physical or intuitive arguments but, for a precise formulation, one must restrict oneself to gauge-invariant quantities.

Let us first consider the Coulomb gauge. In Abelian gauge theories the potentials $A$, or charge-annihilation operators $\phi$, have finite matrix elements, and the bare vacuum has finite energy density. The Coulomb gauge in non-Abelian theories is much more complicated, however. The longitudinal component of the electric field is then given by the formula [for O(3)]

$$E_L=\nabla(\nabla^2\delta^{\alpha\beta}+g\epsilon^{\alpha\gamma\beta}A_k^\gamma\nabla_k)^{-1}\rho\,. \tag{2.1}$$

We note that the denominator is the sum of two terms, one of which is a quantum-mechanical operator. This latter term is absent in the Abelian theory. The Hamiltonian density has the form

$$H=\tfrac{1}{2}E_L{}^2+\text{positive-definite terms}. \tag{2.2}$$

Now the matrix element of the square of the operator on the right of (2.1) will be infinite unless the state is carefully chosen. To see this let us consider the analogous case in particle quantum mechanics. We are then interested in the matrix element

$$\langle (a+bx)^{-2}\rangle , \tag{2.3}$$

where $a$ and $b$ are real numbers, $x$ an operator. The matrix element will be infinite unless the wave function for the state vanishes when $x=-a/b$. Returning to our non-Abelian gauge theory, we notice that the first term on the right of (2.2) is not included in the zero-order Hamiltonian $H_0$. The Coulomb-gauge bare vacuum, which is the eigenstate of $H_0$ with the lowest eigenvalue, is thus not chosen with a view to making the matrix element of $E_L{}^2$ finite. Hence, if we take the expectation value of the complete Hamiltonian density for the Coulomb-gauge bare vacuum, we are likely to obtain an infinite result.

The foregoing argument shows, not that there is an inconsistency in the Coulomb gauge, but that the bare vacuum in this gauge is a bad trial wave function for the true vacuum. In view of this fact, and of the complicated nature of the expression on the right of (2.1), we have not studied the Coulomb gauge further.

The difficulty discussed here is almost certainly related to the difficulties in the Coulomb gauge found by Gribov.[16] His conclusion, like ours, is that the Coulomb gauge in non-Abelian theories is far from straightforward.

Let us now turn to the other commonly used physical gauge, the axial gauge. The bare vacuum in the axial gauge also has infinite energy density, whether the gauge field is Abelian or non-Abelian. This feature was first noticed by Schwinger.[14] For the Abelian theory it is easy to find a "modified bare vacuum" with finite energy density, but an analogous construction does not work for non-Abelian theories.

In the axial gauge, the canonical variables are $A_1^\alpha, A_2^\alpha; E_1^\alpha, E_2^\alpha$. The gauge condition is

$$A_3^\alpha = 0 , \tag{2.4}$$

and $E_3^\alpha$ is found from Gauss's law:

$$E_3^\alpha(x_1,x_2,x_3)=\int_{-\infty}^{x_3} dx_3'\, \mathcal{G}^\alpha(x_1,x_2,x_3') , \tag{2.5a}$$

where

$$\mathcal{G}^\alpha = -\frac{\partial E_1^\alpha}{\partial x_1} - \frac{\partial E_2^\alpha}{\partial x_2} + g\rho . \tag{2.5b}$$

The Hamiltonian is given by the formula

$$H=\tfrac{1}{2}\int d^3x\,[E_3^\alpha(x)]^2 + \text{positive-definite terms.} \tag{2.6}$$

Furthermore, we may write

$$[E^\alpha(x)]^2 = \left| \int \frac{dk_1}{k_1}\, \tilde{\mathcal{G}}(x_1,x_2,k_1)e^{ik_1x}\right|^2 , \tag{2.7a}$$

where

$$\tilde{\mathcal{G}}(x_1,x_2,k)=(2\pi)^{-1/2}\int dx_3 e^{-ikx_3}\mathcal{G}(x_1,x_2,x_3) . \tag{2.7b}$$

We notice from (2.7a) that the expectation value of the Hamiltonian contains an infrared singularity from the small-$k$ region of the integral. In fact

$$\langle \mathcal{H}\rangle = \infty \text{ unless } \frac{1}{k}\mathcal{F}(x_1,x_2,k)\to 0 , \quad k\to 0 \tag{2.8a}$$

where

$$\langle \tilde{\mathcal{G}}(x_1,x_2,k_1)\tilde{\mathcal{G}}(x_1,x_2,k_2)\rangle = \mathcal{F}(x_1,x_2,k_1)\delta(k_1-k_2) . \tag{2.8b}$$

It is easily verified that the condition (2.8) is *not* satisfied in the axial-gauge bare vacuum.

The axial gauge possesses a residual gauge invariance, namely invariance under those gauge transformations which are dependent on $x_1$ and $x_2$ but independent of $x_3$. The generators of such gauge transformations are the operators

$$\begin{aligned} G(x_1,x_2) &\equiv \int_{-\infty}^{\infty} dx_3 \mathcal{G}(x_1,x_2,x_3) \\ &= (2\pi)^{1/2}\tilde{\mathcal{G}}(x_1,x_2,0) . \end{aligned} \tag{2.9}$$

It follows from (2.8) that a necessary condition for finite energy density of the vacuum is

$$\langle G(x_1,x_2)\rangle = 0 . \tag{2.10}$$

In other words, the vacuum must be annihilated by the gauge transformations which are local in $x_1$ and $x_2$, global in $x_3$. Note, however, that (2.10) is not a sufficient condition for the finiteness of this particular term in the Hamiltonian. We require not only the vanishing of $\langle |\mathcal{G}(k)|^2\rangle$ for $k$ precisely equal to zero, but also a sufficiently fast decrease as $k$ approaches zero.

We have emphasized that the energy density of the axial-gauge bare vacuum is infinite in both the Abelian and the non-Abelian theory, and indeed the infinity has a simple physical interpretation. Suppose that one introduces a charge fluctuation in the vacuum at a point $\vec{x}$, but does not change any of the other canonical variables. Since Gauss's law is true as an identity, and since, by hypothesis, we are leaving $E_1$ and $E_2$ unchanged, it follows that an infinitely long tube of electric flux must extend from the point $\vec{x}$ in the 3-direction. The infinite term in the energy density is associated with the infinite length of this flux tube.

The above analysis also indicates how we can cure the problem in Abelian gauge theories. We explicitly introduce the Coulomb fields in $E_1$ and $E_2$ whenever we have a charge fluctuation; in other words, we spread out the flux associated with the fluctuation. In mathematical terms, we define a "modified bare vacuum" in the axial gauge as follows:

$$|0\rangle_{\rm MB}=\exp\left[i\int dx\,dx'\rho(x)\times A_i(x')(x'-x)_i|x'-x|^{-3}\right]|0\rangle_{\rm B}\,. \qquad (2.11)$$

On commuting the operator $E_i(x')$ through the exponential in (2.11), we obtain an extra term

$$\int dx\rho(x)(x'-x)_i|x'-x|^{-3}\,,$$

which is precisely the Coulomb term associated with a charge distribution $\rho(x)$. It is not difficult to check explicitly that the state $|0\rangle_{\rm MB}$, unlike the bare vacuum, does satisfy the condition (2.8) for finite energy density.

Another way of obtaining the state (2.11) is to re-express the Coulomb-gauge bare vacuum in axial gauge.

The foregoing prescription does not work for non-Abelian gauge theories. The electric field itself is then charged, so that the factor corresponding to the exponential in (2.11) not only associates the Coulomb field with original charge $\rho$, but also adds an extra term to the charge density. One may check explicitly that the state (2.11) in a non-Abelian theory does not satisfy the condition (2.8).

We now wish to make a few remarks on the implication of the condition (2.10) with regard to matrix elements. For each value of $x_1$ and $x_2$ we have a different O(3) group of residual gauge transformations. According to (2.10), the vacuum must be an $S$ state with respect to each of these O(3) groups. The vacuum-expectation value of any irreducible-tensor operator, other than a scalar operator, must therefore be zero. For instance,

$$\langle F^{\alpha}_{\mu\nu}(\vec{x}),F^{\alpha}_{\rho\sigma}(\vec{y})\rangle=0\,,\quad \text{unless } x_1=y_1\,,\quad x_2=y_2\,, \qquad (2.12)$$

since the product of the $F$'s is a vector operator with respect to the O(3) groups taken at the points $x_1, y_1$ and the points $x_2, y_2$. To obtain a vacuum-expectation value which is not zero, we must take a path between $\vec{x}$ and $\vec{y}$ and include the line integral which converts (2.12) to a gauge-invariant quantity:

$$\langle F^{\alpha}_{\mu\nu}(\vec{x})V^{\alpha\beta}F_{\rho\sigma}(\vec{y})\rangle\neq 0\,, \qquad (2.13)$$

$$V^{\alpha\beta}=\prod\{\delta^{\alpha_r,\alpha_{r+1}}+\epsilon^{\alpha_r,\beta_r,\alpha_{r+1}}A^{\beta_r}_{\mu}(x_r)[x^{\mu}_{r+1}-x^{\mu}_r]\}\,, \qquad (2.14)$$

$$\alpha_1=\alpha\,,\quad \alpha_{N+1}=\beta\,,$$

where we have divided the path into a large number $N$ of infinitesimal segments bounded by $x_{r+1}$ and $x_r$. The factors in (2.13) are to be ordered along the path.

We can now see that vacuum-expectation values involving potentials must be infinite. Suppose that the coordinates $x_3$ and $y_3$ of $\vec{x}$ and $\vec{y}$ differ by a very small amount, and take a path between $\vec{x}$ and $\vec{y}$ which lies along the 3-axis except for a short portion around a point $\vec{z}$. The vacuum-expectation value (2.12) will be zero, whereas (2.13) will be finite. The only difference between the operators lies in the contribution to $V^{\alpha\beta}$, Eq. (2.14), from the short nonvertical portion of the path near the point $\vec{z}$. It follows that the expectation value

$$\epsilon^{\alpha\beta\gamma}\langle F^{\alpha}_{\mu\nu}(\vec{x})A^{\beta}_i(\vec{z})F^{\gamma}_{\mu\nu}(\vec{y})\rangle \quad (i=1 \text{ or } 2)$$

must be infinite. We stress that none of the points $\vec{x}$, $\vec{y}$, and $\vec{z}$ coincide; this is an infrared and not an ultraviolet infinity.

In the temporal gauge ($A_0=0$) it is also true that expectation values of gauge-dependent quantities are not finite. We now have a residual gauge invariance under all time-independent local gauge transformations. Gauss's law is applied as an extra condition, which requires physical states to be annihilated by the generators of all such gauge transformations. We can invoke arguments similar to those we have used for the axial gauge and can show that vacuum-expectation values of gauge-dependent quantities will not be finite.

Our remarks about matrix elements of gauge-dependent operators in axial and temporal gauges apply equally to Abelian and non-Abelian theories. The difference is that, in Abelian theories, it is possible to find gauges such as the Coulomb gauge where one can define finite potentials. In non-Abelian theories one must restrict oneself to gauge-invariant quantities if one wishes to have finite matrix elements.

We may also mention that the Abelian theory with electric charges and magnetic monopoles is similar to the non-Abelian theory in the features just discussed. It is not evident that one can define a Coulomb gauge; if one tries to do so in the most straightforward way one obtains a vacuum with infinite energy density. The essential difference between the Abelian theory without monopoles on the one hand, and the non-Abelian theory or the Abelian theory with monopoles on the other, is that electric charge is quantized in the latter two theories. One therefore encounters difficulties if one attempts to spread out electric flux in a manner which ignores this quantization.

## III. KINEMATICAL DEFINITION OF GAUGE THEORIES

### A. Operators in gauge theories

In this section we shall examine the kinematic definition of a gauge theory, i.e., a definition in terms of the operators and states of the theory, without reference to a specific Lagrangian. It is only from the viewpoint of such a definition that the theory is symmetric in electric and magnetic quantities.

We begin by reiterating the well-known fact that local gauge invariance is not a *symmetry* in the usual sense of the word. Normally, a symmetry operation takes states of the physical Hilbert space into other states of the physical Hilbert space. One can classify the states of the physical Hilbert space with respect to the representations of the symmetry group. No such classification occurs for physical states with respect to local gauge invariance.

We should rather define a gauge theory in terms of its operator structure. A gauge theory is local in the sense that its $S$ matrix has the usual analytic properties, but it cannot be defined in terms of local, physical operators. One may introduce unphysical local operators such as the potentials or charge-creation operators, but only gauge-invariant operators constructed from them have physical significance. Alternatively one could define the theory entirely in terms of physical, gauge-invariant operators, but local operators such as the product $F^{\alpha}_{\mu\nu}(x)F^{\alpha}_{\rho\sigma}(x)$ would not be sufficient. We would also require nonlocal operators such as the Wilson loop operator, or the path-dependent product of two spatially separated operators with color indices.

### B. Wilson loops and potentials

For a pure gauge theory, the Wilson loops taken at one particular value of the time provide a complete set of commuting, gauge-invariant operators. Later we shall make a remark on the conjugate variables.

The loop operators must satisfy two fundamental relations. The first is that, if a portion of a loop turns back upon itself, one may omit it without changing the value of the operator. Thus, the operator corresponding to the loop in Fig. 1(a) is the same as that in Fig. 1(b).

The second fundamental property of the Wilson loop is that which distinguishes theories with different gauge groups. Let us consider the O(3) Wilson loop

$$\prod_{r=1}^{N}[\delta_{a_r,a_{r+1}}+iL^{\alpha}_{a_r,a_{r+1}}A^{\alpha}_{\mu}(x_r)dx^{\mu}_r]\,, \qquad (3.1)$$
$$\alpha_{N+1}=\alpha_1\,,\quad dx^{\mu}_r=x^{\mu}_{r+1}-x^{\mu}_r\,.$$

The fact that the subscripts $a$ take only two values

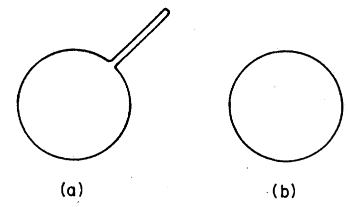


FIG. 1. Two equivalent Wilson loops.

enable us to obtain relations between loops whose perimeters are constructed out of common lines. These relations follow from the identity

$$\delta_{ab}\delta_{cd}\delta_{ef}+\delta_{ad}\delta_{cf}\delta_{eb}+\delta_{af}\delta_{ed}\delta_{cb}-\delta_{ab}\delta_{cf}\delta_{ed}-\delta_{cd}\delta_{af}\delta_{eb}-\delta_{ad}\delta_{cb}\delta_{ef}=0$$
$$(a,b,c,d,e,f=1,2)\,. \qquad (3.2)$$

Applying (3.2) to the loops in Fig. 2(a), we may write

$$4W_1+W_2+W_3-2W_4-2W_5-2W_6=0\,. \qquad (3.3a)$$

This relation may be contrasted with the simpler relation of the Abelian theory [Fig. 2(b)],[17]

$$W_1=W_2\,. \qquad (3.4)$$

We also have the unimodularity condition

$$2W^2-W^{(2)}=1\,, \qquad (3.3b)$$

where $W^{(2)}$ is the loop obtained by going twice around $W_1$. If we had operators which satisfied (3.3a) but not (3.3b), it is not too difficult to show that we could obtain operators satisfying both equations by dividing by $(W^2-W^{(2)})^{1/2}$.

One may also consider loop integrals with $L$ matrices inserted at one point. Such operators are not gauge-invariant and only have meaning as auxiliary quantities. For instance, let $W^{\alpha}_1$, $W^{\beta}_2$, and $W^{\gamma}_3$ represent the auxiliary operators obtained by inserting matrices $2L^{\alpha}$, $2L^{\beta}$, and

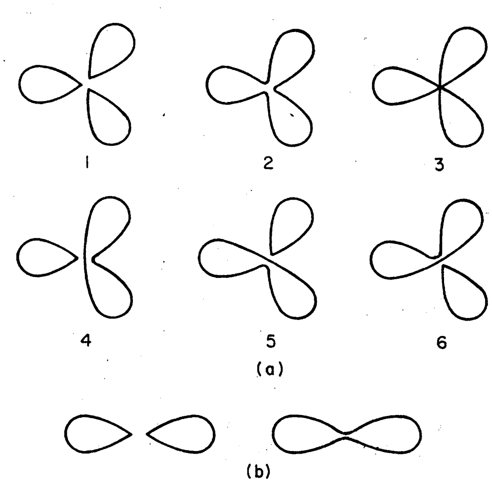


FIG. 2. The loops referred to in Eqs. (3.3) and (3.4).

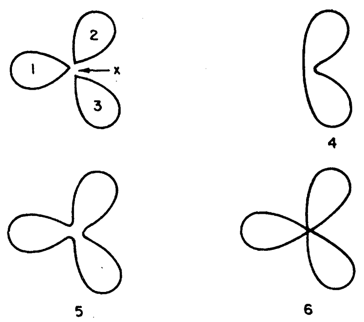


FIG. 3. The loops in the definitions (3.6).

$2L^\gamma$ at the point $x$ in the loops 1, 2, 3 of Fig. 3. Let $W_1, W_2, W_3$ be the corresponding loop operators without the insertions, and let $W_4$, $W_5$, and $W_6$ be the operators associated with the remaining loops of Fig. 3. Then

$$W_2^\alpha W_3^\alpha = W_4 - W_2 W_3\,, \qquad (3.5a)$$

$$\epsilon^{\alpha\beta\gamma} W_1^\alpha W_2^\beta W_3^\gamma = i(W_5 - W_6)\,. \qquad (3.5b)$$

These relations follow if the $W$'s are defined in terms of potentials in the usual way but, from our point of view, the operators $W^\alpha$ have no significance except when inserted into the left-hand side of equations such as (3.5).

Our notation for the auxiliary quantities implies that we are allowed to manipulate them as if the operator $W^\alpha$ could be defined separately, with the superscript $\alpha$ taking three values. As an example, we might have eight loops coming to a point as in Fig. 3. We could then write down the relation

$$\sum(-1)^P \delta^{\alpha\epsilon'} \delta^{\beta\zeta'} \delta^{\gamma\eta'} \delta^{\delta\theta'} W_1^\alpha W_2^\beta W_3^\gamma W_4^\delta W_5^\epsilon W_6^\zeta W_7^\eta W_8^\theta = 0\,. \qquad (3.6)$$

where the indices $\epsilon', \zeta', \eta', \theta'$ represent a rearrangement of $\epsilon, \zeta, \eta, \theta$, and $P$ is the number of permutations required to effect the rearrangement. Similarly, if we have six loops coming to a point, we can write the relation

$$\left(\epsilon^{\alpha\beta\gamma}\epsilon^{\delta\epsilon\zeta} - \sum(-1)^P \delta^{\alpha\delta'} \delta^{\beta\epsilon'} \delta^{\gamma\zeta'}\right) \times W_1^\alpha W_2^\beta W_3^\gamma W_4^\delta W_5^\epsilon W_6^\zeta = 0\,. \qquad (3.7)$$

Equations (3.6) and (3.7), and other such consistency relations, follow from the definitions (3.5), together with the fundamental relation (3.3) applied to any combination of loops.

One can obtain relations similar to (3.5) where the terms carry a net vector index. Thus (Fig. 3)

$$i\epsilon^{\alpha\beta\gamma} W_3^\beta W_2^\gamma + W_3^\alpha W_2 + W_3 W_2^\alpha = W_4^\alpha\,, \qquad (3.8)$$

where $W_4^\alpha$ is defined with the insertion on the right-hand side of the loop 4. As with (3.5), this equation can be obtained by defining the $W$'s in terms of potentials but, from our point of view, they are justified by their consistency with Eqs. (3.5).

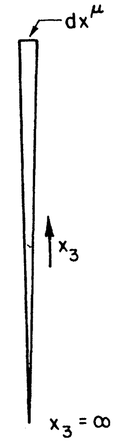


FIG. 4. Wilson loop used to define the axial-gauge potentials.

Once we have defined the loop integral, we can obtain more general matrix elements. For instance, we may write

$$i\{F_{ij}(x)W\}\sigma^{ij} = W\{P + \delta_x P\} - W\{P\}\,, \qquad (3.9)$$

where the expression $W\{P+\delta_x P\}$ denotes the loop with the addition of an infinitesimal area $\sigma$ at the point $x$. In terms of potentials, the operator $\{F_{ij}(x)W\}$ denotes the integral (3.1), together with an insertion $L^\alpha F_{ij}^\alpha(x)$ at the point $x$. We can define a loop with more than one insertion in the same way. If we take a loop with two insertions, and the loop goes back upon itself between the insertions, we obtain the operator in the matrix element on the left of (2.13).

We may define a loop with an electric field insertion by its commutation relations:

$$[\{E_i(x)W_1\}, W_2] = \int_2 dx_i'\,\delta(x - x')W_1^\alpha W_2^\alpha\,. \qquad (3.10)$$

The potentials, like other gauge-dependent operators, are regarded as auxiliary quantities which have no significance except when inserted into gauge-independent expressions. We may obtain the axial-gauge potential $A_i^\alpha(x)$ from the loops by using the formula

$$A_i^\alpha(x)dx^i = -2iW^\alpha\,, \qquad (3.11)$$

where $W^\alpha$ corresponds to the loop shown in Fig. 4; the two long sides extend along the $x_3$ axis to the point $x_3 = -\infty$, and the short side is an element $dx^i$ at the point $x$. The $\alpha$ insertion is at the point $x_3 = -\infty$.

From the potentials we can reconstruct the Wilson integral for a loop of arbitrary shape.

This may be done by repeated application of the formula

$$(W_1\delta_{ab}+2W_1^\alpha L_{ab}^\alpha)(W_2\delta_{bc}+2W_2^\beta L_{bc}^\beta)=W_3\delta_{ac}+2W_3^\gamma L_{ac}^\gamma\,, \tag{3.12}$$

where 1 and 2 refer to the individual loops, and 3 the entire loop, in Fig. 5. Equation (3.12) follows from (3.5) and (3.8). The formula for a loop in terms of potentials is of course the same as Wilson's original formula.

We have thus found that the potentials and the loop integrals form equally good starting points for defining the operators of gauge theories, and it is possible to go from one to the other. The only physically significant operators are those which are gauge invariant. Loop operators with $L$ insertions, or operators corresponding to loops which extend to infinity in the $z$ direction, do not themselves have any real meaning. Matrix elements of such operators between physical (finite-energy) states are not finite.

### C. Color vibrations in electric vortices

In a system with electric confinement, where there exist particles or resonances associated with the Wilson flux loops, Eqs. (3.3) or (3.4) correspond to an important property of the loops, namely the types of (color) spin waves which they can support. We are interested in those color excitations whose energy tends to zero with increasing size of the loop; they could in principle be observed by analyzing the spectrum of high-angular-momentum resonances.

Let us consider the operator we obtain by inserting the factor

$$1+i\epsilon L^\alpha\Phi^\alpha(x) \tag{3.13}$$

at any point $x$ in the loop (3.1), where $\Phi^\alpha$ is any operator of unit color spin. For $\Phi^\alpha$ we might take a local operator such as $F_{\mu\nu}^\alpha(x)$ or, more generally, a smeared-out operator such as

$$\int dx' f(x')V^{\alpha\beta}(x,x')F_{\mu\nu}^\beta(x')\,, \tag{3.14}$$

$V$ being defined over any path joining $x$ and $x'$. If this modified loop operator is applied to the vacuum, the extra energy due to insertion (3.13) will remain finite as the length of the loop increases. On the other hand, it is easily seen by making a perturbation expansion that the change of the state vector increases with the length of the loop and, moreover, that the change is not a simple multiplication of the state vector by a constant. We can conclude that the loop does have excitations whose energy decreases as the length of the loop increases. Such excitations have no analog in the Abelian theory.

To see a direct connection between these excitations and the form of the relation (3.3) or (3.4), we may modify the loop as shown in Fig. 6. For the Abelian case, Eq. (3.4) shows that the addition to the loop operator is simply the product of the operators associated with the individual loops, and the modification will not correspond to any kind of vibrational excitations. For the non-Abelian case the relation is more complicated, and does not lead to any such negative conclusion. It can be seen from (3.5) that the addition to the operator consists of two parts; one is the product of the operators associated with the individual loops, and the other has the form of the second term of (3.13), with $\Phi^\alpha=W_2^\alpha$. We cannot conclude directly from Eqs. (3.5) that color oscillations are associated with this type of modification. We can, however, use such equations to reformulate the theory in terms of potentials, following which we can repeat the reasoning just given.

### D. Gauge field in interaction with charged field

For simplicity we have restricted our analysis in this section to a pure gauge field. We could easily extend our considerations to a gauge field in interaction with other charged fields. If we have a charged field with unit color spin, the fundamental gauge-invariant quantities would be Wilson loops, together with the operators

$$\{\Phi(x)W\} \tag{3.15}$$

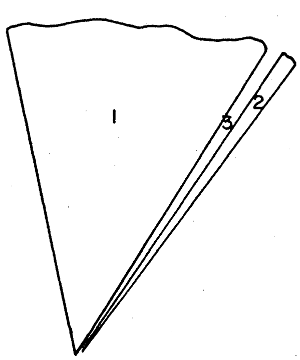


FIG. 5. Reconstruction of the Wilson loop from the potentials.

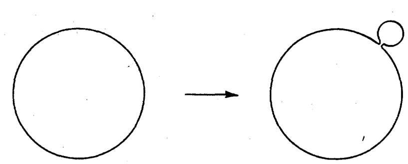

FIG. 6. Modification of a loop with which color vibrations may be associated.

which are defined by making an insertion $\Phi^\alpha(x)L^\alpha$ at the point $x$ in the loop. If we make two such insertions and let the loop go back upon itself between the two point, we obtain the operator

$$\Phi^\alpha(x)V^{\alpha\beta}\Phi^\beta(y)\,, \tag{3.16}$$

where $V^{\alpha\beta}$ is the line integral (2.14). We can then define auxiliary quantities $\Phi^\alpha(x)V^{\alpha\beta}(x,y)$. By extending the path along the $x_3$ axis from the point $x_1, x_2, x_3$ to the point $x_1, x_2, -\infty$, we obtain the axial-gauge $\Phi$'s.

It is important to realize that, if we are simply asking for a set of operators in terms of which all others can be constructed, rather than a set in terms of which the Lagrangian assumes a simple form, the set is by no means uniquely defined. We could add to the potential any gauge-covariant quantity which is a vector in ordinary and color space. In this respect a gauge theory is no different from any other field theory; the set of operators used to define the theory only assumes a certain degree of uniqueness if we demand that the Lagrangian assume a simple form when expressed in terms of those operators.

### E. Abelian loop integrals

Besides the non-Abelian loop integral, one can also define Abelian loop integrals in a non-Abelian theory. They are the loop integrals relevant to the Abelian gauge group which remains after a partial Higgs breaking of the original group. By an Abelian loop integral we mean a loop integral with the simple composition law (3.4) instead of the law (3.3).

An Abelian loop must be defined with respect to an isotopic vector operator $\Phi^\alpha(x)$. For simplicity we assume that $\Phi^\alpha$ commutes with $A_i^\beta$. In the Higgs model such an operator corresponds to an elementary field, but we make no such restriction in our present, general arguments. Indeed, $\Phi$ may be a spatially smeared operator or a product of such operators, always multiplied by line integrals such as $V^{\alpha\beta}(x,x')$ so that it transforms according to the gauge group at $x$. The loop integral would then be quasilocal instead of local; it would fail to commute with other gauge-invariant operators whose field points were within a certain finite distance from the loop itself.[18]

We now define the operator $\Omega$ which rotates $\hat{\Phi}$ into the isotopic 3 direction:

$$\hat{\Phi}\cdot L = \Omega L^3 \Omega^\dagger \quad (\hat{\Phi}^\alpha = \Phi^\alpha/|\Phi|)\,. \tag{3.17}$$

$\Omega$ is undefined to within a factor $\exp[ig\eta(x)L_3]$ on the right. Under a gauge transformation $\Omega$ behaves as follows:

$$\Omega \to (1 - igL\cdot\chi)\Omega\,. \tag{3.18}$$

Following the usual procedure in the Higgs model, we define an Abelian potential by the equation

$$A_\mu^\Phi = \Phi^\alpha A_\mu + 2\frac{i}{g}\,\mathrm{Tr}\,\frac{\partial\Omega}{\partial x^\mu} L_3 \Omega^\dagger\,. \tag{3.19}$$

The Abelian potential remains unchanged if $\Phi$, $A$, and $\Omega$ are subjected to non-Abelian gauge transformations. On the other hand, if we multiply $\Omega$ by a factor $\exp[ig\eta(x)L_3]$ on the right, $A^\Phi$ undergoes a new, Abelian gauge transformation:

$$A^\Phi \to A_\mu^\Phi + \frac{\partial\eta}{\partial x_\mu}\,. \tag{3.20}$$

The Abelian loop operator is defined by the equation

$$W^\Phi = \exp\left(-ig\int dx^i A_i^\Phi(x)\right), \tag{3.21}$$

and it is independent of the function $\eta$.

We observe that the curl of the Abelian potential is just the 't Hooft tensor[2]:

$$\frac{\partial A_\nu^\Phi}{\partial x^\mu} - \frac{\partial A_\mu^\Phi}{\partial x^\nu} = \hat{\Phi}\cdot F_{\mu\nu} - \frac{1}{g}\epsilon^{\alpha\beta\gamma}\hat{\Phi}^\alpha (D_\mu\hat{\Phi}^\beta)(D_\nu\hat{\Phi}^\gamma)$$
$$\equiv F_{\mu\nu}^\Phi\,. \tag{3.22}$$

In the spirit of our present approach we should really define the space components of $F^\Phi$ by (3.22), and the components $F_{0i}^\Phi$ by their commutation relations with $F_{ij}^\Phi$. In the case where $\Phi$ commutes with all components $F_{\mu\nu}^\alpha$, i.e., where it is a separate field, we would have the simple equation

$$E_i^\Phi = \hat{\Phi}\cdot E_i\,. \tag{3.23}$$

An alternative way of defining the Abelian loop is to insert projection operators between all elements of a non-Abelian loop:

$$W^\Phi = \tfrac{1}{2}\,\mathrm{Tr}\prod_r \tfrac{1}{2}[1 + 2L^\alpha\hat{\Phi}^\alpha(x_r)] \times [1 - igL^\alpha A_i^\alpha(x_n)dx^i_{r,r+1}]\,, \tag{3.24}$$

where $x_r$ are a set of points infinitesimally spaced along the loop, and

$$dx_{r,r+1} = x_{r+1} - x_r\,.$$

The easiest way of proving the equivalence of (3.24) and (3.21) is to transform to a gauge where $\Phi$ is in a fixed direction in isospace. With a certain amount of algebra we can also prove directly that the two definitions are equivalent.

## IV. MAGNETIC LOOP OPERATORS

We now wish to define operators which create a loop of magnetic instead of electric flux. The operators must be gauge invariant, since only such operators are physically meaningful. We start with the Abelian magnetic loop, and then proceed to the definition of the non-Abelian loop.

As 't Hooft pointed out, what is required is an operator which is a gauge rotation except on the loop itself. On going once round the loop, the total gauge rotation must be $2\pi$ (or an odd multiple thereof).

### A. Magnetic loop in an Abelian gauge theory

Before dealing with a non-Abelian theory, it will be useful to rewrite, in axial gauge, the fairly standard formula for the construction of a magnetic flux tube in an Abelian theory. We shall make the simplification of taking the tube to be a cylinder whose axis is a straight line in the $x_1$ direction. The radius of the cylinder can be made arbitrarily small at the end of the calculation.[19]

The following classical potential distribution gives us the magnetic field we require, together with an unwanted $\delta$-function contribution at the center of the cylinder which reduces the total flux to zero:

$$A_2=\frac{1}{g}f(\rho)\frac{x_3-x_{30}}{\rho^2}\,,\quad A_3=-\frac{1}{g}f(\rho)\frac{x_2-x_{20}}{\rho^2}\,, \tag{4.1}$$

where

$$\rho^2=(x_2-x_{20})^2+(x_3-x_{30})^2\,, \tag{4.2a}$$

$$f(\rho)=1,\quad \rho=0\,,$$
$$f(\rho)=0,\quad \rho>a\,. \tag{4.2b}$$

We can now subtract out the unwanted singularity at the line $x_2=x_{20}, x_3=x_{30}$ by adding a $\delta$-function contribution to $A$ over a half-plane ending at this line, the direction of $A$ being perpendicular to the plane. If we take the $x_1-x_3$ plane, we obtain

$$A_2=\frac{1}{g}f(\rho)\frac{x_3-x_{30}}{\rho^2}-\frac{2\pi}{g}\delta(x_2-x_{20})\theta(x_3-x_{30})\,, \tag{4.3a}$$

$$A_3=-\frac{1}{g}f(\rho)\frac{x_2-x_{20}}{\rho^2}\,. \tag{4.3b}$$

The operator which creates this potential is

$$\exp\left[-\frac{i}{g}\int d^3x\left(E_2(x)f(\rho)\frac{x_3-x_{30}}{\rho^2}-2\pi E_2(x)\delta(x_2-x_{20})\theta(x_3-x_{30})-E_3(x)f(\rho)\frac{x_2-x_{20}}{\rho^2}\right)\right]. \tag{4.4}$$

Expression (4.4) is gauge invariant. In the axial gauge we may get rid of the apparent singularity at $x_2=x_{20}$ by using the formula (2.5). The expression (4.4) then becomes

$$\exp\left[-\frac{i}{g}\int d^3x\left(E_2(x)f(\rho)\frac{x_3-x_{30}}{\rho^2}-2\pi E_2(x)\delta(x_2-x_{20})\theta(x_3-x_{30})+\mathcal{G}(x)h(x)\right)\right]\,, \tag{4.5a}$$

where

$$h(x_1,x_2,x_3)=\int_{-\infty}^{x_3}dx_3'f(\rho')\frac{x_2-x_{20}}{\rho'^2}\,,\quad \rho'^2=(x_2-x_{20})^2+(x_3'-x_{30})^2\,. \tag{4.5b}$$

Owing to the singularity of the integrand in (4.5b) at $\rho'=0$, the function $h$ changes its value from $-\pi$ to $\pi$ as the variable $x_2-x_{20}$ changes sign. Referring to the expression (2.5b) for $\mathcal{G}$, we observe that the third term has integral eigenvalues and, as far as this term is concerned, a change of $2\pi$ in the function $h$ causes no change in (4.5a). On the other hand, the second term in (2.5b), together with the step-function contribution to $h$, gives a $\delta$-function contribution to $E_2$ from the last term in the exponent of (4.5a). This contribution just cancels the explicit $\delta$-function contribution.

There remains a singularity in (4.5a) along the line $x_1=x_{10}$, $x_2=x_{20}$, since the operator rotates the phase of charged operators by $2\pi$ as we go around this line. This singularity is unimportant for our purposes, since it causes no ambiguity and the whole operator in any case becomes singular in the limit $a\to 0$. If we wish we can remove it by applying (4.5a), not to the actual vacuum, but to the vacuum in the presence of the external potential (4.1).

### B. Abelian magnetic loop in a non-Abelian theory

It is now not difficult to generalize this construction to the Abelian loop in the non-Abelian model. The directions in color space of the operators $E$ have to be specified. As this must be done in a gauge-invariant manner, we select an operator $\Phi$ which serves as our reference. The remarks concerning $\Phi$ which we made when treating the Abelian electric loop apply here too.

We define the gauge-invariant operator

$$E^{\Phi}(x)=E^{\alpha}(x_1,x_2,x_3)V^{\alpha\beta}\hat{\Phi}^{\beta}(x_1,x_{20},x_{30}), \tag{4.6}$$

$V$ being the operator (2.14), taken along the straight-line path linking the points $x_1,x_2,x_3$ and $x_1,x_{20},x_{30}$. The expression (4.4) becomes replaced by

$$M^{\Phi}\equiv \exp\left[-\frac{i}{g}\int d^3x\left(E_2^{\Phi}(x)f(\rho)\frac{x_3-x_{30}}{\rho^2}-2\pi E_2^{\Phi}(x)\delta(x_2-x_{20})\theta(x_3-x_{30})-E_3^{\Phi}(x)f(\rho)\frac{x_2-x_{20}}{\rho^2}\right)\right]. \tag{4.7}$$

Again we can rewrite the expression in the axial-gauge form analogous to (4.5), but the details are slightly more complicated. We require the operator $\mathcal{G}^{\Phi}(x_1,x_2,x_3,x_3')$, defined

$$\mathcal{G}^{\Phi}(x_1,x_2,x_3,x_3')=\mathcal{G}^{\alpha}(x_1,x_2,x_3)V^{\alpha\beta}(x_1,x_2,x_3';x_1,x_{20},x_{30})\Phi^{\beta}(x_1,x_{20},x_{30}), \tag{4.8}$$

where $V$ is the line integral (2.14), taken along the straight-line path between its arguments. The operator (4.8) is invariant with respect to transformations in the residual gauge group. We can then replace (4.7) by the expression

$$M^{\Phi}\equiv \exp\left[-\frac{i}{g}\int d^3x\left(E_2^{\Phi}(x)f(\rho)\frac{x_3-x_{30}}{\rho^2}-2\pi E_2^{\Phi}(x)\delta(x_2-x_{20})\theta(x_3-x_{30})\right.\right.$$
$$\left.\left.-\int_{-\infty}^{x_3}dx_3'\,\mathcal{G}^{\Phi}(x_1,x_2,x_3,x_3')f(\rho')\frac{x_2-x_{20}}{\rho'^2}\right)\right]. \tag{4.9}$$

As in the Abelian case, we can now show that there is in fact no singular contribution to (4.9) when $x_2=x_{20}$. For this value of $x_2$ the dependence on $x_3'$ of $\mathcal{G}^{\Phi}$ disappears, since the factor $V^{\alpha\beta}$ in (4.8) becomes equal to unity. The last term in the exponent of (4.9) thus adopts a form similar to that of the Abelian case:

$$-\frac{i}{g}\int d^3x\,\mathcal{G}^{\Phi}(x)h(x_2,x_3)=-\frac{i}{g}\int d^3x\,\hat{\Phi}^{\alpha}(x_1,x_{20},x_{30})\left(-\frac{\partial E_1^{\alpha}}{\partial x_1}-\frac{\partial E_2^{\alpha}}{\partial x_2}+g\rho(x)\right)h(x_2,x_3). \tag{4.10}$$

Again the function $h$ changes its value from $-\pi$ to $\pi$ as $x_2-x_{20}$ changes sign, so that the second term in the large parentheses of (4.10) has a $\delta$-function contribution which cancels the second term in the exponent of (4.9).

In the present case, the first term in the large parentheses of (4.10) gives a nonzero contribution if $\hat{\Phi}^{\alpha}$ depends on $x_1$. This contribution *appears* to be multiplied by a step function of $x_2-x_{20}$; if it were the operator (4.9) would increase $A_1$ by a step function $(x_2-x_{20})$, and would thus be singular at $x_2=x_{20}$. The step-function contribution is, however, illusory, as may be seen by writing

$$\exp[i(A+B)]=\exp\left(i\int_0^1 d\eta e^{i\eta B}Ae^{-i\eta B}\right)\exp(iB), \tag{4.11}$$

with

$$A=-\frac{1}{g}\int d^3x\left(\frac{\partial}{\partial x_1}\hat{\Phi}^{\alpha}(x_1,x_{20},x_{30})\right)E_1^{\alpha}(x)h(x_2,x_3),$$

$$B=-\int d^3x\,\hat{\Phi}^{\alpha}(x)h(x_2,x_3).$$

When $h$ changes from $\pi$ to $-\pi$ the effect of the change of the factors $e^{\pm i\eta B}$ in (4.11) is to rotate the factor $E_1^{\alpha}$ in $A$ by $\pi$ about the $\hat{\Phi}^{\alpha}$ axis. Since $\hat{\Phi}^{\alpha}\hat{\Phi}^{\alpha}=1$, $\hat{\Phi}^{\alpha}\,\partial\hat{\Phi}^{\alpha}/\partial x_1=0$, the effect of this rotation is to change the sign of $A$ and thus to compensate the sign change caused by the factor $h$ itself in $A$.[20] We thus conclude that the expression on the right of (4.9) has no singularity when $x_2=x_{20}$.

As in the Abelian case, we can now conclude from (4.7) that, apart from a possible singular contribution when $x_2=x_{20}$, the operator $M^{\Phi}$ effects no physical change outside the region $f(\rho)\neq 0$, and that it is independent of the direction chosen for the $x_3$ axis in the axial gauge. The expression (4.9) shows that the possible singular contribution at $x_2=x_{20}$ is in fact absent.

Thus far we have been considering "loops" which are straight lines along the $x_1$ axis. In the general case, one can again define the loop operator by (4.7), but taken with respect to a curvilinear coordinate system. The distance $x_1$ is marked off along the loop, planes of fixed $x_1$ are taken perpendicular to the loop, and, in each plane, a rectangular $x_2-x_3$ coordinate system is taken. For future reference it is convenient to define the $x_2-x_3$ integral in the exponent of (4.7) by the symbol $-(i/g)\mathcal{B}^{\Phi}(x_1)$, so that

$$M^{\Phi}=\exp\left(-\frac{i}{g}\int dx_1\mathcal{B}^{\Phi}(x_1)\right). \tag{4.12}$$

We should emphasize that the expression on the right of (4.12) makes sense only if we integrate $x_1$ over a closed loop (or an infinitely long line). The arguments following (4.5) or (4.10) could not be carried through if the $x_1$ integral were to terminate; an extra singularity would be introduced at the terminating value of $x_2$ and *all* values of $x_3$. We could have anticipated this restriction, since a magnetic flux line of unit strength can only terminate on a Dirac monopole.

### C. Non-Abelian magnetic loop

Though the Abelian magnetic loop operators do have the properties required by 't Hooft, they are not sufficient to construct all operators of the system. They are equivalent to one rather than three sets of potentials $A_i$ per point. Furthermore, it is known that physical magnetic vortices in non-Abelian theories require two operators to specify their direction in color space. The magnetic flux can have any direction relative to these two operators, and the loop can support color oscillations. The Abelian magnetic loops cannot support such oscillations. They are analogous to the Abelian electric loop operators discussed in the previous section; what we require is a magnetic operator analogous to the non-Abelian Wilson loop operator.

We may express the Wilson operator as a sum of contributions as follows:

$$W=\sum_{a_1,\ldots,a_N}\prod_r[\delta_{a_r,a_{r+1}}+igL^{\alpha}_{a_r,a_{r+1}}A^{\alpha}_i(x_r)dx^i_{r,r+1}]\,. \qquad (4.13)$$

The coordinates $x_r$ and the differentials have the same meaning as in (3.1). The color spinor indices $a_r$ take the values 1, 2. Equation (4.13) may be reexpressed in the form

$$W=\sum_{a_1,\ldots,a_N}\prod_r W_{a_r,a_{r-1}}(x_1)\,, \qquad (4.14a)$$

where

$$W_{ab}(x_r)=\tfrac{1}{2}(\delta_{ab}+1)\exp(igL^{\alpha}_{ab}A^{\alpha}_i\,dx^i_{r,r+1})$$
$$+\tfrac{1}{2}(\delta_{ab}-1)\exp(-igL^{\alpha}_{ab}A^{\alpha}_i\,dx^i_{r,r+1})\,. \qquad (4.14b)$$

Even before we pass to the limit $dx^i_{r,r+1}\to 0$, each element $W_{ab}$ creates half a unit of flux along the direction $dx^i_{r,r+1}$. Only in the limit $dx^i_{r,r+1}$ does it behave correctly under gauge transformations.

We can now construct the magnetic non-Abelian loop in a similar way. In order to obtain a gauge-invariant operator, we shall define our coordinate system *internally*. We take three operators $\hat{\Phi}^{\alpha}_1, \hat{\Phi}^{\alpha}_2, \hat{\Phi}^{\alpha}_3$ of unit color spin; the operators are normalized and mutually orthogonal in color space. We then define components $\mathcal{B}_{\alpha}$ with respect to these cases:

$$\mathcal{B}_{\alpha}\equiv\mathcal{B}^{\Phi_{\alpha}}\,. \qquad (4.15)$$

Analogously to (4.14), we define

$$M=\sum_{a_1,\ldots,a_N}\prod_r M_{a_r,a_{r+1}}(x_r)\,, \qquad (4.16)$$

where

$$M_{ab}(x_r)=\tfrac{1}{2}(\delta_{ab}+L^{\alpha}_{ab})\exp\left(\frac{i}{g}\mathcal{B}_{\alpha,i}\,dx^i_{r,r+1}\right)$$
$$+\tfrac{1}{2}(\delta_{ab}-L^{\alpha}_{ab})\exp\left(-\frac{i}{g}\mathcal{B}_{\alpha,i}\,dx^i_{r,r+1}\right)\,. \qquad (4.17)$$

For each combination of subscripts $a_1,\ldots,a_N$, we have a magnetic loop of the form (4.7). At those junctions of elements $dx^i_{r,r+1}$ where the color direction of the vector $E^{\alpha}$ changes suddenly, it is implied that this change is smoothed out over a distance small compared to $dx^i_{r,r+1}$.

The expression (4.17) does not satisfy (3.3b) but, as we have explained, we can now modify the definition so that it does.

It is certainly not evident that our formulas tend to a definite limit as the loop becomes infinitely thin or the segments in (4.17) become infinitely short (J. Polchinski, private communication). To obviate this difficulty, we construct our magnetic variables on a lattice; the thickness of the loops is of the order of the lattice dimensions. As the lattice may be made arbitrarily fine and as we are not attempting to answer dynamical questions, this is adequate for our purpose.

The expression (4.16) is unaltered if the $L^{\alpha}$'s in (4.17) are subjected to an O(3) rotation. In this way, we obtain a *magnetic* gauge group, which must be distinguished from the electric gauge group.

We can modify the above definitions slightly so that two-loop operators (taken at a particular time) always commute. The product of any number of loop operators, applied to the vacuum, is defined with the product of the exponentials in (4.17) replaced by the exponential of the sum of the exponents. Having defined such a product, we can define inductively the state $M|n\rangle$, where $|n\rangle$ denotes a product of $n$-loop operators applied to the vacuum. Two-loop operators so constructed will obviously commute.

### D. Magnetic variables

From our method of constructing the non-Abelian magnetic loop, it is clear that the relation (3.3) will be satisfied. We can therefore construct magnetic vector potentials in the same way as we constructed the electric vector potentials. The loops, or other gauge-invariant operators, can be constructed from them using the standard formulas, and the totality of these operators comprises all the operators of the theory. However, to write the Hamiltonian in terms of the magnetic vector potentials would in practice be very complicated.

It is important to realize that the electric and magnetic variables are not related by any simple equation of the form

$$\tilde{F}^{\alpha,\,\mu\nu} = \epsilon^{\mu\nu\rho\sigma} F_{\alpha,\,\rho\sigma}\,. \tag{4.18}$$

The color indices on the two types of operators do not correspond. It is true that

$$\tilde{F}^{\alpha,\,\mu\nu}(x)\,N^{\alpha\beta}(x,y)\,\tilde{F}^{\beta,\,\mu\nu}(y) \approx \epsilon^{\mu\nu\rho\sigma}\epsilon^{\mu\nu\lambda\tau}F^{\alpha}_{\rho\sigma}(x)V^{\alpha\beta}(x,y)F^{\beta}_{\lambda\tau}(y)\,,$$

$$(x-y)^2 \to 0;\ \text{no summation over } \mu \text{ and } \nu\,, \tag{4.19}$$

where $V$ is defined by (2.14), and $N$ is its magnetic counterpart. Unfortunately, we cannot conclude from (4.19) that the Hamiltonian in the electric and magnetic variables are the same, since the Hamiltonian really depends on the terms of the expressions in (4.8) which are linear in $(x-y)$.

Given any electric operator, one can construct a corresponding magnetic operator. Analogously to (3.15), one can define

$$\{\Psi_M(x)M\} \tag{4.20}$$

as the insertion of a factor

$$\Psi^{\alpha}(x)\hat{\Phi}^{\alpha}_{\gamma}L^{\gamma}$$

at a point $x$ in the formula (4.16). According to this prescription, the operator

$$\{\hat{\Phi}_{\gamma,\,M}(x),M\}\,, \tag{4.21}$$

where $\hat{\Phi}_{\gamma}$ is one of the operators used to specify the internal coordinate system, is defined by simply inserting a factor $L^{\gamma}$ at the point $x$.

In our construction of the magnetic loop operator from electric variables, we began with the Abelian loop and then proceeded to the non-Abelian loop. Now that we have defined our magnetic variables, we should be able to use the analogs of the electric formulas to reconstruct the magnetic Abelian loop. Let us construct the Abelian loop associated with one of the operators $\hat{\Phi}^{\alpha}_{\gamma}$ which we used to specify our coordinate system. It is easiest to use the analog of the formula (3.2) and, according to the prescription just given, the factor $1+2L^{\alpha}\hat{\Phi}^{\alpha}(x_n)$ simply becomes $1+2L^{\gamma}$. An insertion of this factor between each factor of (4.16) leads us back to our original Abelian loop with $\Phi=\Phi_{\gamma}$; this may most easily be seen by taking $\gamma=3$, in which case the factor $1+2L^{\gamma}$ restricts all subscripts $a_n$ to the value 1. According to (4.17),

$$M_{11}(x_n) = \exp\left(-\frac{i}{g}\,\mathcal{B}_{3,\,i}\,dx^{i}_{n,\,n+1}\right) \equiv \exp\left(-\frac{i}{g}\,\mathcal{B}^{\Phi_3}dx^{i}_{n,\,n+1}\right). \tag{4.22}$$

The result then follows from (4.12).

We have seen that the 't Hooft tensor in the electric representation can be obtained by taking a small Abelian electric loop. We could similary define the space components of a magnetic 't Hooft tensor by taking a small magnetic loop. It is simplest to take the loop in a plane containing the $x_3$ axis. From (4.7), we observe that the only term in the exponent which survives in the limit $f(\rho)=0$, $\rho\neq 0$, is the $\delta$-function term. The magnetic 't Hooft tensor is thus simply

$$F^{\Phi_3,\,M}_{ij} = \epsilon^{ijk}\,\hat{\Phi}^{\alpha}_{3}E^{\alpha}_{k}\,. \tag{4.23}$$

Equation (4.22) is only true as it stands if $\Phi^{\alpha}_{3}$ is a separate field which commutes with all components $F^{\alpha}_{\mu\nu}$ since, in the more general case, Eq. (4.7) itself must be modified. When $\hat{\Phi}^{\alpha}_{3}$ is a separate field, we note from (3.2) and (4.22) that

$$F^{\Phi_3,\,M}_{ij} = \epsilon^{ijk}E^{\Phi_3}_{k}\,. \tag{4.24}$$

Hence, while the lack of correspondence between color indices prevented us from writing down an equality such as (4.18), we observe that a similar equality *is* true for the 't Hooft tensor. In the more general case, the two sides of (4.24) may differ by terms involving the commutator of $\Phi$ with the gauge fields.

## V. PHASES OF NON-ABELIAN GAUGE THEORIES

In this section we wish to use the foregoing analysis to make certain remarks about the phases of non-Abelian gauge theories. Four phases may be distinguished [in the O(3) theory]:

(i) The phase with real massless gluons.

(ii) The phase with partial Higgs symmetry breaking and a remaining U(1) invariance group.

(iii) The phase with complete Higgs symmetry breaking.

(iv) The confinement phase.

To our knowledge it has never been proved that we cannot have complete Higgs symmetry breaking *and* confinement; we feel that such a phase is most unlikely and we shall not discuss it further.

We begin by remarking that phases (ii), (iii), and (iv) cannot be distinguished from phase (i) by the breaking of a symmetry in the strict sense of the word. In Abelian or non-Abelian theories we should not consider local gauge invariance as a genuine symmetry; in non-Abelian gauge theories even global gauge invariance should not be so considered. The generators $\tilde{G}^\alpha$ of these transformations $[\tilde{G}^\alpha = \int dx_1 dx_2 G^\alpha(x_1, x_2)]$, or the quantities $\tilde{G}^\alpha \tilde{G}^\alpha$, are not invariant with respect to gauge transformations local in $x$ and $y$, global in $z$. They therefore do not correspond to physically observable quantities; their matrix elements between physical states are not finite. One cannot use the generators to define quantum numbers. One might attempt to define color quantum numbers by the electric flux at large distances from the object in question, but this quantity, too, is not gauge invariant. Non-Abelian theories hereby differ from Abelian theories, where charge and electric flux are gauge-invariant operators.

We can also infer in a less formal manner that global non-Abelian gauge invariance does not constitute a symmetry in the usual sense, since there exists no phase in which one expects to observe definite multiplets. In particular, in phase (i), an arbitrary small disturbance causes the emission of an infinite number of soft colored gluons, and it would seem to be impossible to determine the color of a given object.

Witten[21] has stressed that we also cannot distinguish between phases by the nonvanishing of a vacuum-expectation value. The operator $\Phi^\alpha$ is not gauge invariant, and we therefore have to consider an operator such as

$$\Phi^\alpha(x) V^{\alpha\beta}(x, -\infty), \qquad (5.1)$$

where $V$ is the line integral (2.14), taken along any path between $x$ and $-\infty$. The vacuum-expectation value of (5.1) is zero even in the Higgs phase, since the factor $V$ provides a contribution of the form $e^{-L}$, with $L$ infinite. In fact, the vanishing of expection values of axial-gauge $\Phi$'s is a particular case of this effect; the path is then taken to $-\infty$ along the $x_3$ axis.

In Abelian theories without monopoles we can characterize the Higgs phase by the nonvanishing of the vacuum-expectation value of the Coulomb-gauge $\Phi$, i.e., of the gauge-invariant operator

$$\Phi(x) \exp\left(ie \int d^3x' \frac{(x'-x)_i A_i(x')}{|x'-x|^3}\right). \qquad (5.2)$$

In non-Abelian theories the Coulomb-gauge operators probably have no meaning, and we cannot construct a gauge-invariant operator from a formula similar to (5.2). In Abelian theories with monopoles, too, the operator (5.2) has no physical significance, since one may easily verify that the total energy which it creates diverges at large distances. In fact, the Higgs procedure, as modified by Bardakci and Samuel,[22] enables us to reformulate a non-Abelian theory as an Abelian theory with monopoles; in either case one cannot distinguish between the phases by the vanishing or otherwise of vacuum-expectation values.

### A. Phase with real massless gluons

Let us now examine the four phases in turn. Whether phase (i) can exist in a non-Abelian theory and, if so, whether it can be interpreted physically, are unanswered questions. We suspect that the answers to both are negative, but the complicated nature of the phase in question has prevented us from obtaining a proof. We know that Nature does not choose this particular phase; any theory which may have (i) as the phase of lowest energy can immediately be rejected on experimental grounds. At the present time it is certainly not known how to construct a trial vacuum state corresponding to the phase (i) in a non-Abelian theory whereas, in an Abelian theory, one can simply take the Coulomb-gauge bare vacuum. One *might* remove the infinite energy density order by order in perturbation theory and hope that the procedure makes sense in the limit of infinite order, even if the perturbation series itself diverges. It is highly unlikely that this hope would be realized in an infra-red unstable theory.

We should stress that the above remarks are meant to apply to the zero-temperature vacuum. Polyakov and Susskind have suggested that a phase analogous to (i) may be realized at high temperatures. Our arguments have no bearing on the possible existence of such a phase, whose properties would be completely different from those of the zero-temperature phase.

With zero-mass gluons we could construct states of arbitrary low mass with any color quantum number, in so far as such a quantum number can be defined. We could also construct states of arbitrary low mass containing Yang-Wu monopoles in any combination. Phase (i), if it can exist, thus appears to be symmetric in charges and monopoles though, owing to our lack of understanding of this phase, we hesitate to make such a statement with certainty.

### B. Phase with unbroken U(1) invariance group

Phase (ii) is characterized by the presence of Abelian massless vector particles, and therefore by a gauge-invariant conserved charge. The divergenceless current may be defined as follows:

$$\tilde{j}^\mu = \hat{\Phi}^\alpha j^{\alpha\mu} + \frac{\partial \hat{\Phi}^\alpha}{\partial x^\nu} F^{\alpha,\mu\nu}, \qquad (5.3)$$

where $j^{\alpha\mu}$ is the non-gauge-invariant color current. The total charge associated with this current will be

$$\epsilon^{ijk}\int d\sigma_{ij}\hat{\Phi}^{\alpha}F^{\alpha}_{k0}\,, \qquad (5.4)$$

the integral to be taken over a fixed-time closed surface at large distances from the system in question. In phase (ii), the vacuum is an eigenstate of this charge.

Phase (ii) also has a conserved quantum number associated with 't Hooft-Polyakov monopoles. Arafune, Freund, and Goebel[23] have shown that this is simply the number of zeroes in $\Phi$, a zero at $\vec{x}_0$ being taken positive or negative according as, after a continuous transformation, $\Phi^{\alpha,i}(x) = \pm\epsilon^{\alpha ij}(x-x_0)$. The number of monopoles is equal to the integral, over a distant fixed-time surface, of the magnetic flux associated with the 't Hooft tensor

$$\int d\sigma^{ij}F^{\Phi}_{ij}\,, \qquad (5.5)$$

where $F^{\Phi}$ is defined in (3.22).

Though the charge-current density (5.3), and the positions of the zeroes in $\Phi$, are dependent on the actual choice of $\Phi$, the total electric charge (5.4) and magnetic charge (5.5) are independent of this choice, except for an overall normalization constant. In the absence of any massless particles, the integrals of any vector or axial-vector field operator, taken over surfaces distant from any matter, will vanish. If there exists a single massless vector particle, as is the case with phase (ii), the integrals will be proportional to the conserved (electric or magnetic) charge with which the vector particle interacts. Apart from the proportionality constant, they will be independent of the precise vector or axial-vector field operator chosen. It may, of course, happen that the constant of proportionality is zero.

We have already noted that passage from the electric to the magnetic variables interchanges the components $F^{\Phi}_{ij}$ and $\epsilon^{ijk}E^{\Phi}_{k}$ of the 't Hooft tensor in the special case where $\Phi$ commutes with all the fields $F^{\alpha}_{\mu\nu}$. In the general case the relation between the 't Hooft tensors in the two representations is more complicated, but it is unlikely that the constant of proportionality between the fluxes $F^{\Phi,M}_{ij}$ or $E^{\Phi,M}_{i}$ and the electric or magnetic charges would be reduced to zero. Hence, when we change from electric to magnetic variables, the electric and magnetic charges become interchanged, and the phase as a whole is symmetric in electric and magnetic variables.

### C. Phases with magnetic or electric confinement

We now turn to the discussion of phases (iii) and (iv). 't Hooft[10] has shown that, in the absence of massless particles either the electric loop $W$ or the magnetic loop $M$ must behave like the exponential of the area if the loop becomes large, so that the phases enumerated above are the only ones possible (provided we exclude simultaneous electric and magnetic confinement). We shall examine phase (iv); owing to the symmetry between electric and magnetic quantities a precisely analogous discussion could be made for phase (iii). We remind the reader that our whole treatment refers to the case where no actual quark fields are present.

The definition of confinement as the absence of color nonsinglets, while extremely useful intuitively, can certainly not be made precise, since we cannot make a gauge-invariant definition of a color singlet. In no phase would we expect to see color multiplets. We also pointed out that we could not distinguish between phases by the vanishing or nonvanishing of vacuum-expectation values. Most recent work on confinement defines the phase by one of the three properties:

(a) $W\sim e^{-A}$ for a large loop,

(b) If two external quark (i.e., color spinor) sources are introduced, the energy increases linearly with their separation,

(c) The system can support quasistable color vortices with half a unit of flux [for O(3)]. These vortices can shrink to a point and then disintegrate, or they can form a figure-of-eight and split into two, but they cannot dissipate their energy by diffusion.

't Hooft[10] proposed using definitions similar to (a) or (c) for phase (iii). It was already known from the work of Nielsen and Olesen that the phase with complete Higgs symmetry breaking could support quasistable magnetic vortices with half a unit of flux.

Our treatment of definition (a) differs slightly from that of Wilson in that we are examining loops at a fixed time. Within this framework we wish to make it plausible that definitions (a) and (b) are equivalent. Our analysis will indicate how we might apply the Wilson condition to construct trial state vectors for confined hadronic states. A very similar argument could be made for the equivalence of the definitions (a) and (c), the only difference being that we take a closed loop of flux instead of a tube stretched between the two quark sources.

The question to be answered is whether the half unit of electric flux between the two sources can lose energy by spreading out. A gauge-invariant state with the two quark sources could be defined in the usual way as

$$Q_a W_{ab}\{\vec{x}(x_3)\}Q_b|0\rangle\,, \qquad (5.6)$$

where $W$ is the spinor line-integral operator (similar to the Wilson loop before taking the trace), and $\vec{x}(x_3)$ $[\equiv x_1(x_3), x_2(x_3)]$ represents a path between the two quarks whose coordinates will be taken to be $0, 0, \pm L$.

The state (5.6) has an infinite energy, since the flux tube stretching between the two quarks has zero thickness. It is a nontrivial problem to spread out flux in a non-Abelian gauge theory (or in an Abelian theory with monopoles). One way would be to take a functional integration over different paths between the quarks:

$$\int \mathfrak{D}\{x(x_3)\} f\{\vec{x}(x_3)\} Q_a W_{ab}\{\vec{x}(x_3)\} Q_b |0\rangle , \qquad (5.7a)$$

where

$$\int d\vec{x} f\{\vec{x}(x_3)\} = 1, \quad -L < x_3 < L , \qquad (5.7b)$$

$$f\{\vec{x}(x_3)\} \to \delta^2(\vec{x}), \quad x_3 \to \pm L . \qquad (5.7c)$$

Actually, we have simplified the notation in writing down (5.7), since we should allow paths which go backwards in the $x_3$ direction for part of their length. We must also allow closed loops of flux in the integration, without which it would not be possible to get rid of the divergence in question. We shall assume that it is possible to define a state of the form (5.7) with finite energy density (apart from the vacuum energy density) away from the quarks themselves.

Now let us investigate the behavior of the energy as the distance between the quarks is increased. We are interested in the value of the quantity

$$\frac{\int d^3x'' \int \mathfrak{D}\{\vec{x}(x_3)\}\mathfrak{D}\{\vec{x}'(x_3)\} f\{\vec{x}(x_3)\} f\{\vec{x}'(x_3)\}\langle Q_b^\dagger W_{ba}\{\vec{x}'(x_3)\} Q_a^\dagger \{E_i^\alpha(x'')\}^2 Q_c W_{cd}\{\vec{x}(x_3)\} Q_d\rangle}{\int \mathfrak{D}\{\vec{x}(x_3)\}\mathfrak{D}\{\vec{x}'(x_3)\} f\{\vec{x}(x_3)\} f\{\vec{x}'(x_3)\}\langle Q_b^\dagger W_{ba}\{\vec{x}'(x_3)\} Q_a^\dagger Q_c W_{cd}\{\vec{x}(x_3)\} Q_d\rangle} . \qquad (5.8)$$

The functional integrand in the denominator of (5.8) consists of a closed connected or disconnected Wilson loop, together with insertions $Q_a^\dagger Q_c, Q_d Q_b^\dagger$ at the quark positions. The numerator also consists of such loops, together with an extra factor $[E_i^\alpha(x'')]^2$. As the flux is spread out, the behavior of the function integral will thus depend entirely on that of large Wilson loops; the extra quark insertions or $E^2$ factors will not effect the dependence of such loops on the length or the area.

First, let us suppose that we are in a phase other than (iv), so that

$$\langle W\rangle = e^{-cL/l} , \qquad (5.9)$$

$L$ being the length of the loop and $l$ its thickness (after spreading out the flux). If we omit a region surrounding the quarks in the $x''$ integral and then scale all dimensions (including the omitted region) in the expression (5.8) by a given factor, the value of the expression does not change. Hence, by spreading out the flux as we separate the quarks, we can prevent the energy from increasing indefinitely, and we have no confinement.

Now let us suppose that

$$\langle W\rangle \approx e^{-\alpha A} . \qquad (5.10)$$

As a first orientation, we shall evaluate (5.8) in lowest-order perturbation theory in the $E_i$'s, but exactly in the Wilson loop operators. In other words, we consider disconnected diagrams in which each factor $E_i$ is paired with one factor $A_i$ from the loops. Furthermore, let us suppose that one of the $A_i$ factors is from the bra, one from the ket. The contribution of this term to the numerator of (5.8) is as follows:

$$\tfrac{1}{2}\int d^3x'' \int \mathfrak{D}\{\vec{x}(x_3)\}\mathfrak{D}\{\vec{x}'(x_3)\} f\{\vec{x}(x_3)\} f\{\vec{x}'(x_3)\}\, \delta^2(\vec{x}'' - \vec{x}(x_3''))\, \delta^2(\vec{x}'' - \vec{x}'(x_3''))\left(1 + \frac{\partial \vec{x}(x_3'')}{\partial x_3''} \cdot \frac{\partial \vec{x}'(x_3'')}{\partial x_3''}\right)$$
$$\times \langle Q_b^\dagger W_{ba}^\alpha\{\vec{x}'(x_3)\} Q_a^\dagger Q_a W_{ab}^\alpha\{\vec{x}(x_3)\} Q_b\rangle . \qquad (5.11)$$

The superscript $\alpha$ on the $W$'s denote $L^\alpha$ insertions at the point $x_3''$.

At a particular value $x_{30}''$ of $x_3''$, the $\delta$ functions in (5.11) restrict the Wilson loops to those where $\vec{x}(x_{30}'') = \vec{x}'(x_{30}'')$. There is no such restriction in the denominator of (5.8). In the nonconfinement case where (5.9) holds, the values of the loop integrals do not fall off appreciably as $\vec{x}(x_3'')$ and $\vec{x}'(x_3'')$ are separated. The range of integration of $\vec{x}(x_3'')$ and $\vec{x}'(x_3'')$, for values of $x_3''$ near $x_{30}''$, becomes much larger in the denominator than in the numerator as the flux is spread out $[f\{x\} \neq 0$ for a large range of values of $x]$. The energy per unit length of $x_3$ thus decreases.

If, on the other hand, $\langle W\rangle \sim e^{-\alpha A}$, the denominator receives no appreciable contribution from values of $\vec{x}(x_3'')$ and $\vec{x}'(x_3'')$ which are well separated, since a separation of $\vec{x}(x_3'')$ and $\vec{x}'(x_3'')$ will increases the area of the loop. The important range of integration of $\vec{x}(x_{30}'') - \vec{x}'(x_{30}'')$ in the denominator remains finite as the flux is spread out, the energy per unit length between the quarks does not decrease, and we have confinement.

In the foregoing argument we have concentrated on a particular term in the perturbation series. In general, the $x''$ integrand in (5.8) would differ appreciably from the vacuum-expectation value $\int d^3x'' \langle \{E_i^\alpha(x'')\}^2\rangle$ only if the point $x''$ is near one or more points $\vec{x}'(x_3), x_3$ or $\vec{x}''(x_3), x_3$ in the loop. For large loops, one may distinguish contributions where the point $x''$ is near at least one point $\vec{x}'(x_3), x_3$ from the bra *and* one point $\vec{x}(x_3), x_3$ from the ket, and contributions where $x''$ is near points from the bra or the ket alone. The argument given above can be extended to general contributions of the former type.

If the cluster involving the $E$'s contains points exclusively within the bra or the ket, it would appear that we had confinement independently of the behavior of $\langle W\rangle$. According to our scaling argument, this would mean that we could not remove the divergence in (5.6) by spreading out the flux in the manner described. As we have pointed out above, we may well be able to overcome the difficulty by extending our functional integral to include closed loops of flux. Let us assume that we can do so. If we then scaled all dimensions upward, the Wilson criterion would decrease the effect of the closed loops, and the energy per unit length of $x_3$ would eventually cease to decrease. Again, therefore, we would have confinement.

Basically, the closed loops of flux, either in the ket or bra themselves or in the overlap integral, prevent a thick flux tube from being a simple classical superposition of thinner tubes. As long as the large loops are ineffective we cannot decrease the energy by spreading out the flux.

If a tube of flux between two distant quarks cannot lose its energy by spreading out, it follows that a large closed tube cannot lose its energy by diffusion. A system in phase (iv) can thus support closed "strings" of flux. We have already pointed out that the types of color oscillation modes of the strings will distinguish between theories with different gauge groups. The situation is precisely analogous to that of the quasistable closed Nielsen-Olesen vortices which can be supported by a system in phase (iii).

In our entire discussion, we have assumed that the only way of spreading out flux in a gauge-invariant manner is to integrate functionally over flux lines of different shape. We know of no other way of spreading out flux in a system where flux is quantized, i.e., a system where we cannot simply create a flux line of arbitrary strength. In continuum QED without monopoles, on the other hand, one can easily spread out flux by integrating within the exponential,[24] to obtain a state of the form

$$\int d\lambda Q^\dagger \exp\left(-ie\int_P dx^i A_i(x)\right) Q\,|0\rangle\,, \qquad (5.12)$$

where $\lambda$ is a parameter defining the path $P$. Spreading out the flux within the exponential decreases the energy independently of the Wilson criterion. It thus appears that we cannot have confinement in continuum QED without monopoles. This result appears to be in contradiction with Wilson's original argument, which indicates that his criterion would lead to confinement in Abelian or non-Abelian gauge theories. The only possible resolution between the contradictory results appears to be the hypothesis that the Wilson criterion cannot be satisfied in continuum QED without monopoles.

One might ask which phase the Weinberg-Salam model chooses. Neither monopoles nor Nielsen-Olesen vortices are present, and the criteria suggested above would not identify either phases (ii) or (iii). In fact, there appears to be no obvious way in which a strongly coupled Weinberg-Salam model would differ from an ordinary Abelian gauge theory.

We should like to suggest that an effect occurs similar to that in the $(2+1)$-dimensional Georgi-Glashow model, which has been studied by Polyakov.[25] Both models possess instantons centered around a zero in the Higgs field; the classical Weinberg-Salam instantons are zero-size, finite-energy objects, but 't Hooft[26] has shown that quantum effects would give them a finite size. In the Georgi-Glashow model the instantons change the Higgs phase into a confined phase. In the Weinberg-Salam model, too, we would expect the instantons to remove the Higgs effect; as they are unlikely to bring back the massless gluons, the only possibility is, again, a confined phase. [In the actual weak-coupled Weinberg-Salam model such instanton effects, dependent on $\exp(-1/g^2)$, would give a confinement distance several orders of magnitude greater than the radius of the universe.] The Weinberg-Salam model possesses quark fields, and the confined phase thereof falls outside the scope of the present paper; as Susskind[27] and others have emphasized, we should *not* expect the Wilson criterion to hold.

We may note that Fradkin and Shenker,[28] working from a result of Seiler and Osterwalder,[29] have suggested on entirely different grounds that a

Weinberg-Salam lattice model may confine for certain ranges of the parameters.

We conclude with a brief note on the possible construction of a vacuum in phase (iv) from a vacuum in another phase. An ordinary superconducting vacuum is distinguished from a normal vacuum by the property

$$\langle\phi\rangle \neq 0 \quad \text{(superconducting)}, \tag{5.13a}$$

$$\langle\phi\rangle = 0 \quad \text{(normal)}. \tag{5.13b}$$

A superconducting vacuum can then be constructed from a normal vacuum by forming a coherent plasma of charged objects. 't Hooft has emphasized the analogy between (5.13) and the properties

$$\langle M\rangle \approx e^{-L} \quad \text{(phase iv)} \tag{5.14a}$$

$$\langle M\rangle \approx e^{-A} \quad \text{(phase iii)}. \tag{5.14b}$$

For large loops, the behavior $e^{-L}$ would be expected if there were no special properties which caused the expectation value to vanish. 't Hooft[10] therefore suggested that $e^{-L}$ and $e^{-A}$ in (5.14) should be associated with a nonvanishing or a vanishing expectation value, respectively, in (5.13). It would then follow that, *if we were given a vacuum in phase* (iv), we could construct a confinement vacuum by forming a coherent phasma of Nielsen-Olesen vortices.

The problem which actually faces us in non-Abelian gauge theories is to construct a confinement vacuum from a vacuum in phase (i). More precisely, we wish to construct a vacuum with finite energy density (apart from ultraviolet divergences) and to prove that it is in phase (iv). Since the vacuum from which we start already satisfies the condition (5.14a), we would not expect to obtain a confinement vacuum by forming a plasma of Nielsen-Olesen vortices.

Our initial vacuum does not satisfy the condition (2.10); in fact, our whole problem is to construct a vacuum which does satisfy this criterion, or, more precisely, the stronger criterion (2.8). We have emphasized throughout this paper that the condition (2.10) implies that physical quantities must be invariant under the residual gauge group. Until we have a vacuum which satisfies this criterion it will be necessary to introduce non-gauge-invariant quantities, such as creation operators $\Omega$ for Wu-Yang monopoles. We should then expect that

$$\langle\Omega\rangle \neq 0 \quad \text{(phase iv)}, \tag{5.15a}$$

$$\langle\Omega\rangle = 0 \quad \text{(phase i)}, \tag{5.15b}$$

though we emphasize again that (5.15) cannot be regarded as a precise statement, $\Omega$ not being gauge invariant. For non-Abelian gauge theories, (5.13) is a similar imprecise statement; nevertheless, we can construct a superconducting vacuum by forming a plasma of charged objects. It may therefore be possible to construct a confinement vacuum by taking a plasma of Wu-Yang monopoles. In a later paper we hope to show that the finite-energy criterion (2.8) can be satisfied in such a vacuum.

In the interior of the Nielsen-Olesen vortices the vacuum-expectation value of $\phi$, and therefore the density of the coherent plasma of charged objects, falls to zero. As Nielsen and Olesen stressed, their vortices were the analog of Landau-Ginsburg vortices in type-II superconductors, namely a normal (non-Higgs) phase within the vortex and a superconducting (Higgs) phase outside. By the electric-magnetic analogy, the vortices of electric flux may be viewed as a region of nonconfining phase where the density of Wu-Yang monopoles falls to zero. Depending on whether we have a flux tube between quarks or a closed vortex, the nonconfining region will or will not contain quarks. If the angular momentum of the vortex is high, the contrifugal force will cause the length of the vortex to be much greater than its thickness but, for low-angular-momentum states, the vortex structure may be lost, and we may have an object more like the MIT bag.

## ACKNOWLEDGMENTS

I should like to acknowledge helpful discussions with K. Bardakci and J. Polchinski.

This research was supported by the National Science Foundation under grant PHY77-23512.

[1]T. T. Wu and C. N. Yang, in *Properties of Matter under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York, 1969), p. 349.

[2]G. 't Hooft, Nucl. Phys. B79, 276 (1976).

[3]A. M. Polyakov, Zh. Eksp. Teor. Fiz. Pis'ma Red. 20, 430 (1974) [JETP Lett. 20, 194 (1974)].

[4]H. B. Nielsen and P. Olesen, Nucl. Phys. B61, 45 (1973).

[5]Y. Nambu, Phys. Rev. D 10, 4262 (1974).

[6]S. Mandelstam, Phys. Lett. 53B, 476 (1975).

[7]G. 't Hooft, CERN report (unpublished); J. Kogut and L. Susskind, Phys. Rev. D 11, 395 (1975).

[8]S. Mandelstam, Phys. Rep. 23C, 245 (1976).
[9]G. 't Hooft, in *High Energy Physics*, Proceedings of the European Physical Society International Conference, Palermo, 1975, edited by A. Zichichi (Editrice Compositori, Bologna, 1976), p. 1225.
[10]G. 't Hooft, Nucl. Phys. B138, 1 (1978).
[11]To avoid confusion, we state that we use the term magnetic vector potential to apply to the usual vector potentials, which, in our present terminology, are electric variables. The *electric* vector potentials are the corresponding magnetic variables.
[12]Formal solutions to this problem have in fact been given. Thus far, however, no one has been able to calculate expectation values for operators in a finite-energy trial vacuum.
[13]S. Mandelstam, Bull. Am. Phys. Soc. 22, 541 (1977).
[14]J. Schwinger, Phys. Rev. 130, 402 (1963).
[15]P. Goddard, J. Nuyts, and D. Olive, Nucl. Phys. B125, 1 (1977).
[16]V. N. Gribov, Leningrad report (unpublished).
[17]The Abelian and O(3) loops may also be distinguished by their topological properties; Abelian loops have a direction, while O(3) loops do not. Two O(3) loops are topologically equivalent to none. We do not wish to rely on topological distinctions, since the topology only distinguishes between the center of the gauge group. For instance, the loops in all $O(2n+1)$ theories have the same topology.
[18]We are of course referring to spacelike loops.
[19]For the Abelian theory one could work entirely in terms of a flux tube of zero radius. For the non-Abelian theory such a procedure would give rise to an ambiguity, and it is necessary to take the limit of a tube of finite radius.
[20]This is the point at which it becomes necessary to consider a tube of flux with finite radius, which eventually approaches zero, rather than to take a zero-radius tube from the beginning.
[21]E. Witten, private communication.
[22]K. Bardakci and S. Samuel, Phys. Rev. D 18, 2849 (1978).
[23]J. Arafune, P. G. O. Freund, and C. G. Goebel, J. Math. Phys. 16, 433 (1975).
[24]This point has been emphasized by M. Weinstein, private communication.
[25]A. M. Polyakov, Nucl. Phys. B120, 429 (1977).
[26]G. 't Hooft, Phys. Rev. D 14, 3432 (1976).
[27]L. Susskind, private communication.
[28]E. Fradkin and S. Shenker, SLAC report (unpublished).
[29]K. Osterwalder and E. Seiler, Ann. Phys. (N. Y.) 110, 440 (1978).