WHICH TOPOLOGICAL FEATURES OF A GAUGE THEORY CAN BE RESPONSIBLE FOR PERMANENT CONFINEMENT? *)

G. 't Hooft

Institute for Theoretical Physics
Utrecht, The Netherlands

Lecture II

II1. INTRODUCTION

In the previous lecture a simple gauge model was considered with a scalar field doublet ξ. Perturbation expansion was considered not about the point $\xi = o$ but about the "vacuum value"

$$\xi = \begin{pmatrix} F \\ o \end{pmatrix} .$$

Such a theory is usually called a theory with "spontaneous symmetry breakdown" [1]. In contrast one might consider "unbroken gauge theories" where perturbation expansion is only performed about a symmetric "vacuum". These theories are characterized by the absence of a mass term for the gauge vector bosons in the Lagrangian. The physical consequences of that are quite serious. The propagators now have their poles at $k^2 = o$ and it will often happen that in the diagrams new divergences arise because such poles tend to coincide. These are fundamental infrared divergencies that imply a blow-up of the interactions at large distance scales. Often they make it nearly impossible to understand what the stable particle states are.

A particular example of such a system is "Quantum Chromodynamics", an unbroken gauge theory with gauge group SU(3), and in addition some fermions in the 3-representation of the group, called "quarks". We will investigate the possibility that these quarks are permanently confined inside bound structures that do not carry gauge quantum numbers. First of all this idea is not as absurd as it may seem. The converse would be equally difficult to understand. Gauge quantum numbers are a priori only defined up to local gauge transformations. The existence of *global* quantum

*) Lecture given at the Cargèse Summer Institute, August 26 - September 8, 1979.

numbers that would correspond to these local ones but would be detectable experimentally from a distance is not at all a pre-requisite. We are nevertheless accustomed to attaching a global significance to local gauge transformation properties because we are familiar with the theories with spontaneous breakdown. The electron and its neutrino, for example, are usually said to form a gauge doublet, to be subjected to local gauge transformations. But actually these words are not properly used. Even the words "spontaneous breakdown" are formally not correct for local gauge theories (which is why I put them between quotation marks). The vacuum *never* breaks local gauge invariance because it itself is gauge invariant. All states in the physical Hilbert space are gauge-invariant. This may be confusing so let me illustrate what I mean by considering the familiar Weinberg-Salam-Ward model. The invariant Lagrangian is

$$\mathcal{L}^{inv} = -\tfrac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} - \tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - D_\mu\phi^* D_\mu\phi - V(|\phi|)$$

$$- \bar{\psi}_L\gamma D\psi_L - \bar{e}_R\gamma De_R - \kappa\bar{e}_R(\phi^*\psi_L) - \kappa(\bar{\psi}_L\phi)e_R \quad . \qquad \text{(II1)}$$

Here ϕ is the scalar Higgs doublet. The gauge group is $SU(2)\times U(1)$, to which correspond A^a_μ $(G^a_{\mu\nu})$ and A^o_μ $(F_{\mu\nu})$. The subscripts L and R denote left and right handed components of a Dirac field, obtained by the projection operators $\frac{1}{2}(1\pm\gamma_5)$.

e_R is a singlet;

ψ_L is a doublet.

D_μ stands for covariant derivative.

The function $V(|\phi|)$ takes its minimum at $|\phi| = F$. Usually one takes

$$<\phi>_{vacuum} = \begin{pmatrix} F \\ o \end{pmatrix} \qquad \text{(II2)}$$

and perturbes around that value: $\phi = \begin{pmatrix} F+\tilde{\phi}_1 \\ \tilde{\phi}_2 \end{pmatrix}$.

One identifies the components of ψ with neutrino and electron:

$$\psi = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \quad . \qquad \text{(II3)}$$

However, this model is *not* fundamentally different from a model with "permanent confinement". One could interpret the same physical particles as being all gauge singlets, bound states of the fundamental fields with extremely strong confining forces, due to the gauge fields A_μ^a of the group SU(2). We have scalar quarks (the Higgs field ϕ) and fermionic quarks (the ψ_L field) both as fundamental doublets. Let us call them q. Then there are "mesons" ($q\bar{q}$) and "baryons" (qq). The neutrino is a "meson". Its field is the composite, SU(2)-invariant

$$\phi^*\psi_L = F\nu_L + \text{negligible higher order terms.}$$

The e_L field is a "baryon", created by the SU(2)-invariant

$$\varepsilon_{ij}\phi_i\psi_j = Fe_L + \ldots \quad \text{(II4)}$$

the e_R field remains an SU(2) singlet.
Also bound states with angular momentum occur: The neutral intermediate vector boson is the "meson"

$$\phi^* D_\mu \phi = \frac{i}{2} g F^2 A_\mu^{(3)} + \text{total derivative + higher orders,} \quad \text{(II5)}$$

if we split off the total derivative term (which corresponds to a spin-zero Higgs particle).
The $W_\mu^\pm$ are obtained from the "baryons" $\varepsilon_{ij}\phi_i D_\mu \phi_j$, and the Higgs particle can also be ontained from $\phi^*\phi$.
Apparently some mesonic and baryonic bound states survive perturbation expansion, most do not (only those containing a Higgs "quark" may survive).

Is there no fundamental difference then between a theory with spontaneous breakdown and a theory with confinement? Sometimes there is. In the above example the Higgs field was a faithful representation of SU(2). This is why the above procedure worked. But suppose that all scalar fields present were invariant under the center Z(N) of the gauge group SU(N), but some fermion fields were not. Then there are clearly two possibilites. The gauge symmetry is "broken" if physical objects exist that transform non-trivially under Z_N, such as the fundamental fermions. We call this the Higgs phase. If on the other hand all physical objects are invariant under Z_N, such as the mesons and the baryons, then we have permanent confinement.

Quantum Chromodynamics is such a theory where these distinct possibilities exist. It is unlikely that one will ever prove from first principles that permanent confinement takes place, simply because one can always imagine the Higgs mode to occur. If no fundamental scalar fields exist then one could introduce composite fields such as

$$\phi_{ab} = G_{\mu\nu}^a G_{\mu\nu}^b$$

or

$$\phi_i^j = \bar{\psi}_i \psi^j$$

and postulate nonvanishing vacuum expectation values for them:

$$\langle \phi_{ab} \rangle = F_1 \, d_{ab8} + F_2 \, d_{ab3} \quad \text{or}$$

$$\phi_i^j = F_1 \, \lambda_{8i}{}^j + F_2 \, \lambda_{3i}{}^j \; .$$

In that case there would be no confinement. Whether or not $F_{1,2}$ are equal to zero will depend on details of the dynamics. Therefore, dynamics must be an ingredient of the confinement mechanism, not only topological arguments. What we will attempt in this lecture is to show that topological arguments imply for this theory the existence of phase regions, separated by sharp phase transition boundaries (usually of first order). One region corresponds to what is usually called "spontaneous breakdown", and will be referred to as Higgs phase. Another corresponds to absolute quark confinement. Still another phase exists which allows for long range Coulomb-like forces to occur. (Coulomb phase.)

II2. VORTICES IN THE PERIODIC BOX

We concentrate on long-range topological phenomena. One topological feature is the instanton, corresponding to a gauge field configuration with non-trivial Pontryagin or Second Chern Class number. This however has no direct implication for confinement. What is needed for confinement is something with the space-time structure of a string, i.e. a two dimensional manifold in 4 dim. space-time. Instantons are rather event-like, i.e. zero dimensional and can for instance give rise to new types of interactions that violate otherwise apparent symmetries. We will not consider these further here*. A topological structure which is extended in two dimensional sheets exists in gauge theories, as has been first observed by Nielsen, Olesen [2)] and Zumino [3)]. They are crucial. We will exhibit them by compactifying space-time. For the instanton it had been convenient to compactify space-time to a sphere $S^{(4)}$. For our purposes a hypertorus

$$S^{(1)} \times S^{(1)} \times S^{(1)} \times S^{(1)} \, ,$$

is more suitable [4)]. One can also consider this to be a four dimensional cubic box with periodic boundary conditions. Inside, space-time is flat. The box may be arbitrarily large. To be ex-

---

*Surely, as is explained by C. Callan in his lectures, configurations such as the instanton gas will influence the *dynamics* and thereby give rise to a perhaps crucial force that causes the system to choose the confinement rather than the Higgs mode. The absoluteness of the confining force is however not explained that way.

plicit we put a pure SU(N) gauge theory in the box (no quarks yet). Now in the continuum theory the gauge fields themselves are representations of SU(N)/Z(N), where Z(N) is the center of the group SU(N):

$$Z(N) = \left\{ e^{2\pi i n/N} I \ ; \ n = o,1,\ldots,N-1 \right\} . \tag{II6}$$

This is because any gauge transformation of the type (II6) leaves $A_\mu(x)$ invariant. A consequence of this is the existence of another class of topological quantum numbers in this box besides the familiar Pontryagin number. Consider the most general possible periodic boundary condition for $A_\mu(x)$ in the box. Take first a plane (x_1,x_2) in the 12 direction with fixed values of x_3 and x_4. One may have

$$\begin{aligned} A_\mu(a_1,x_2) &= \Omega_1(x_2)\, A_\mu(o,x_2) \ , \\ A_\mu(x_1,a_2) &= \Omega_2(x_1)\, A_\mu(x_1,o) \ . \end{aligned} \tag{II7}$$

Here, a_1, a_2 are the periods.
ΩA_μ stands short for $\Omega A_\mu \Omega^{-1} + \frac{1}{gi} \Omega \partial_\mu \Omega^{-1}$.
The periodicity conditions for $\Omega_{1,2}(x)$ follow by considering (II7) at the corners of the box:

$$\Omega_1(a_2)\, \Omega_2(o) = \Omega_2(a_1)\, \Omega_1(o) Z \ , \tag{II8}$$

where Z is some element of Z(N).
One may now perform continuous gauge transformations on $A_\mu(x)$,

$$A_\mu(x_1,x_2) \to \Omega(x_1,x_2)\, A_\mu(x_1,x_2) \ , \tag{II9}$$

where $\Omega(x_1,x_2)$ (non-periodic) can be arranged either such that $\Omega_2(x_1) = I$ or such that $\Omega_1(x_2) = I$, but not both, because Z in (II8) remains invariant under (II9) as one can easily verify. We call this element Z(1,2) because the 12 plane was chosen. By continuity Z(1,2) cannot depend on x_3 or x_4. For each $(\mu\nu)$ direction such a Z element exist, to be labeled by integers

$$n_{\mu\nu} = -n_{\nu\mu} \ , \tag{II10}$$

defined modulo N. Clearly this gives

$$N^{\frac{d(d-1)}{2}} = N^6 \tag{II11}$$

topological classes of gauge field configurations. Note that these classes disappear if a field in the fundamental representation of SU(N) is added to the system (these fields would make unacceptable jumps at the boundary). Indeed, to understand quark confinement it is necessary to understand pure gauge systems without quarks first. As we shall see, the new topological classes will imply the existence of new vacuum parameters besides the well-known instanton angle θ.

II3. ORDER AND DISORDER LOOP INTEGRALS

To elucidate the physical significance of the topological numbers $n_{\mu\nu}$ we first concentrate on gauge field theory in a three dimensional periodic box with time running from $-\infty$ to ∞. To be specific we will choose the temporal gauge,

$$A_4 = o \tag{II12}$$

(this is the gauge in which rotation towards Euclidean space is particularly elegant). Space has the topology $S(1)^3$. There is an infinite set of homotopy classes of closed oriented curves C in this space: C may wind any number of times in each of the three principal directions. For each curve C at each time t there is a quantum mechanical operator A(C,t) defined by

$$A(C,t) = \mathrm{Tr}\, P \exp \oint_C ig\vec{A}(\vec{x},t).d\vec{x} \ , \tag{II13}$$

called Wilson loop or order parameter.
Here P stands for path ordering of the factors $\vec{A}(\vec{x},t)$ when the exponents are expanded. The ordering is done with respect to the matrix indices. The $A(\vec{x},t)$ are also operators in Hilbert space, but for different $\vec{x}$, same t, all $A(\vec{x},t)$ commute with each other. By analogy with ordinary electromagnetism we say that A(C) *measures* magnetic flux *through* C, and in the same time *creates* an electric flux line *along* C. Since A(C) is gauge-invariant under purely periodic gauge transformations, our versions of magnetic and electric flux are gauge-invariant. Therefore they are not directly linked to the gauge *co*variant curl $G^a_{\mu\nu}(x)$.

There exists a dual analogon of A(C) which will be called B(C) or disorder loop operator [5]. C is again a closed oriented curve in $S(1)^3$. A simple definition of B(C) could be made by postulating its equal-time communication rules with A(C):

$$[A(C), A(C')] = o \ ;$$

$$[B(C), B(C')] = o \ ; \tag{II14}$$

$$A(C)\, B(C') = B(C')\, A(C) \exp 2\pi i n/N \ ,$$

where n is the number of times C' winds around C in a certain direction. Note that n is only well defined if either C or C' is in the trivial homotopy class (that is, can be shrunk to a point by continuous deformations). Therefore, if C' is in a nontrivial class we must choose C to be in a trivial class. Since these commutation rules (II14) determine B(C) only up to factors that commute with A and B, we could make further requirements, for instance that B(C) be a unitary operator.

An explicit definition of B(C) can be given as follows. In the temporal gauge, $A_o = o$, one must distinguish a "large Hilbert space" $\mathcal{H}$ of all field configurations $\vec{A}(\vec{x})$ from a "physical Hilbert space $H \subset \mathcal{H}$. This H is defined to be the subspace of $\mathcal{H}$ of all gauge invariant states:

$$H = \left\{ |\psi\rangle, \langle \vec{A}(\vec{x})|\psi\rangle = \langle \Omega\vec{A}(\vec{x})|\psi\rangle \right\} \qquad \text{(II15)}$$

where Ω is any infinitesimal gauge transformation in 3 dim. space. Often we will also write Ω for the corresponding rotation in $\mathcal{H}$:

$$H = \left\{ |\psi\rangle, \Omega|\psi\rangle = |\psi\rangle, \Omega \text{ infinitesimal} \right\} . \qquad \text{(II16)}$$

Now consider a pseudo-gauge transformation $\Omega^{[C']}$ defined to be a genuine gauge transformation at all points $x \notin C'$, but singular on C'. For any closed path $x(\theta)$ with $o \leqslant \theta < 2\pi$ twisting n times around C' we require

$$\Omega^{[C']}(x(2\pi)) = \Omega^{[C']}(x(o))\, e^{2\pi i n/N} . \qquad \text{(II17)}$$

This discontinuity is not felt by the fields $A(\vec{x},t)$ which are invariant under Z(N). They do feel the singularity at C' however. We define B(C') as

$$\Omega^{[C']}$$

but with the singularity at C' smoothened; this corresponds to some form of regularization, and implies that the operator differs from an ordinary gauge transformation. Therefore, even for $|\psi\rangle \in H$ we have

$$B(C')\,|\psi\rangle \neq |\psi\rangle . \qquad \text{(II18)}$$

For any regular gauge transformation Ω we have an Ω' such that

$$\Omega\Omega^{[C']} = \Omega^{[C']}\,\Omega' . \qquad \text{(II19)}$$

Therefore, if $|\psi\rangle \in H$ then $B(C')\,|\psi\rangle \in H$, and B(C') is gauge-invariant. We say that B(C') *measures* electric flux *through* C' and *creates* a magnetic flux line *along* C'.

II4. NON-ABELIAN GAUGE-INVARIANT MAGNETIC FLUX IN THE BOX

We now want to find a conserved variety of Non-Abelian gauge-invariant magnetic flux in the 3-direction in the 3 dimensional periodic box. One might be tempted to look for some curve C enclosing the box in the 12 direction so that A(C) measures the flux through the box. That turns out not to work because such a flux is not guaranteed to be conserved. It is better to consider a curve C' in the 3-direction winding over the torus exactly once:

$$C' = \left\{\vec{x}(s)\,,\ o \leq s < 1\ ;\ \vec{x}(1) = \vec{x}(o) + (o,o,a_3)\right\} . \qquad \text{(II20)}$$

B(C') creates one magnetic flux line. But B(C') also changes the number n_{12} into $n_{12}+1$. This is because

$$\Omega^{[C']}$$

makes a Z(N) jump according to (II17). If $\Omega_{1,2}(x)$ in (II7) are still defined to be continuous then Z in (II8) changes by one unit. Clearly, n_{12} measures the number of times an operator of the type B(C') has acted, i.e. the number of magnetic flux lines created. n_{12} is also conserved by continuity. We simply define

$$n_{ij} = \varepsilon_{ijk}\, m_k\ , \qquad \text{(II21)}$$

with m_k the total magnetic flux in the k-direction. Note that $\vec{m}$ corresponds to the usual magnetic flux (apart from a numerical constant) in the Abelian case. Here, $\vec{m}$ is only defined as an integer modulo N.

II5. NON-ABELIAN GAUGE-INVARIANT ELECTRIC FLUX IN THE BOX

As in the magnetic case, there exists no simple curve C such that the total electric flux through C, measured by B(C), corresponds to a conserved total flux through the box. We consider a curve C winding once over the torus in the 3-direction and consider the electric flux creation operator A(C). But first we must study a new conserved quantum number.

Let $|\psi\rangle$ be a state in the before mentioned little Hilbert space H. Then, according to eq. (II16), $|\psi\rangle$ is invariant under *infinitesimal* gauge transformations Ω. But we also have some non-trivial homotopy classes of gauge transformations Ω. These are the pseudo-periodic ones:

$$\begin{aligned} \Omega(a_1,x_2,x_3) &= \Omega(o,x_2,x_3)Z_1\ , \\ \Omega(x_1,a_2,x_3) &= \Omega(x_1,o,x_3)Z_2\ , \\ \Omega(x_1,x_2,a_3) &= \Omega(x_1,x_2,o)Z_3\ , \\ Z_{1,2,3} &\in \text{center } Z(N) \text{ of } SU(N)\ . \end{aligned} \qquad \text{(II22)}$$

Notice that not only do $A_\mu(x)$ transform smoothly under these Ω (they are invariant under Z(N) transformations), but their boundary conditions do not change. These Ω therefore commute with the magnetic flux $\vec{m}$. If two different Ω satisfy the same equation (II22) they act differently on states of the big Hilbert space $\mathcal{H}$, but since they only differ by regular gauge transformations* they act identically on states in H, defined in II16. Thus, $Z_{1,2,3}$, characterized by three integers k_1, k_2, k_3:

$$Z_t = e^{ik_t 2\pi/N} , \tag{II23}$$

define a set of N^3 different operators in H, under which the Hamiltonian H is invariant. Let us call these operators $\Omega[\vec{k}]$. Their eigenstates satisfy

$$\Omega[\vec{k}]|\psi\rangle = e^{i\omega(\vec{k})}|\psi\rangle , \tag{II24}$$

where $\omega(\vec{k})$ are strictly conserved numbers. Since $\Omega[\vec{k}]$ form a (finite) group $(Z(N))^3$ we have

$$\omega(\vec{k}_1) + \omega(\vec{k}_2) = \omega(\vec{k}_1+\vec{k}_2 \text{ mod } N) \text{ mod } 2\pi , \tag{II25}$$

and

$$\omega(o) = o , \tag{II26}$$

because $|\psi\rangle \in H$.

Therefore,

$$\omega(\vec{k}) = \frac{2\pi}{N} \sum_i e_i k_i \ (\text{mod } 2\pi) , \tag{II27}$$

where e_i are three fixed integers, defined modulo N. They are three conserved numbers, to be compared with the instanton angle θ.

Now let us turn back to A(C), defined in eq. (II13). If C is the curve considered in the beginning of this section, A(C) is not invariant under $\Omega[\vec{k}]$, because

---

*Here we left aside for sake of simplicity the fourth integer, the Pontryagin number of the gauge transformations. It gives interesting complications of which Witten's result for magnetic monopoles is an example[6]. To avoid these complications we must restrict ourselves to θ (the vacuum angle associated with instantons) $= o$, a subspace of H. Clearly, generalization to $\theta \neq o$ is possible.

$$A(C) \to \mathrm{Tr}\, \Omega(\vec{x}_1)\, P(\exp \int_c ig\vec{A}\, d\vec{x})\, \Omega^{-1}(\vec{x}_1+\vec{a}_3)$$
$$= e^{-2\pi i k_3/N} A(C) \; . \tag{II28}$$

Therefore,

$$A(C)\, \Omega[\vec{k}]|\psi\rangle = \Omega[\vec{k}]\, e^{-2\pi i k_3/N} A(C)|\psi\rangle \; . \tag{II29}$$

If $\Omega[\vec{k}]|\psi\rangle = e^{i\omega(\vec{k})}|\psi\rangle$, (II30)

and $A(C)\ |\psi\rangle = |\psi'\rangle$, (II31)

then $\Omega[\vec{k}]|\psi'\rangle = e^{i\omega(\vec{k})+2\pi i k_3/N}|\psi'\rangle$. (II32)

Therefore A(C) increases e_3 by one unit:

$$e_3\, A(C)|\psi\rangle = A(C)\ (e_3+1)|\psi\rangle \; . \tag{II33}$$

e_3 is a good indicator for electric flux in the 3-direction. It is strictly conserved. The physical interpretation of the three integers e_i (mod N) is electric flux. It is gauge-invariant and conserved. Notice that neither electric nor magnetic flux can be properly defined if fields in the fundamental representation are present.

II6. THE FREE ENERGY OF A GIVEN FLUX CONFIGURATION AT LOW BUT FINITE TEMPERATURE

The free energy F of a system with given flux quantum $(\vec{m},\vec{e})$ at temperature $T = 1/k\beta$ is given by

$$e^{-\beta F} = \mathrm{Tr}_{H}\, P_e(\vec{e})\, P_m(\vec{m})\, e^{-\beta H} \; . \tag{II34}$$

Here H is the Hamiltonian, and H is the little Hilbert space. P are projection operators. $P_m(\vec{m})$ is simply defined to select a given set of $n_{ij} = \varepsilon_{ijk}\, m_k$, the three space-like indices of eq. (II10). How is $P_e(\vec{e})$ defined? We must select states $|\psi\rangle$ with

$$\Omega[\vec{k}]|\psi\rangle = e^{\frac{2\pi i}{N}(\vec{k}\vec{e})}|\psi\rangle \; , \tag{II35}$$

therefore

$$P_e(\vec{e}) = \frac{1}{N^3} \sum_{\vec{k}} e^{-\frac{2\pi i}{N}(\vec{k}\vec{e})} \Omega(\vec{k}) \ . \tag{II36}$$

Now $e^{-\beta H}$ is the evolution operator in imaginary time direction at interval β, expressed by a functional integral over a Euclidean box with sides (a_1, a_2, a_3, β):

$$<\vec{A}_{(1)}(\vec{x})\ e^{-\beta H}\ \vec{A}_{(2)}(\vec{x})> = \int DA\ e^{S(A)} \Bigg|_{\vec{A}(\vec{x},o) = A_{(2)}(\vec{x})}^{\vec{A}(\vec{x},\beta) = A_{(1)}(\vec{x})} \tag{II37}$$

We may fix the gauge for $A_2(\vec{x})$ for instance by choosing

$$A_{(2)3}(\vec{x}) = o,$$
$$A_{(2)2}(x,y,o) = o\ , \tag{II38}$$
$$A_{(2)1}(x,o,o) = o\ .$$

We already had $A_4(\vec{x},t) = o$. Since only states in H are considered, we insert also a projection operator

$$\int_{\Omega \in I} D\Omega\ ,$$

where I is the trivial homotopy class*.

"Trace" means that we integrate over all $A_{(1)} = A_{(2)}$, therefore we get periodic boundary conditions in the 4-direction. Insertions of $\int_{\Omega\in I} D\Omega$ means that we have periodicity up to gauge transformations, in the completely unique gauge

$$A_4(\vec{x},\beta) = A_3(\vec{x},o) = A_2(x,y,o,o) = A_1(x,o,o,o) = o\ . \tag{II39}$$

Eq. (II36) tells us that we have to consider twisted boundary conditions in the 41, 42, 43 directions and Fourier transform:

$$e^{-\beta F(\vec{e},\vec{m},\vec{a},\beta)} = \frac{1}{N^3} \sum_{\vec{k}} e^{-\frac{2\pi i}{N}(\vec{k}\vec{e})} W\{\vec{k},\vec{m},a\mu\}\ . \tag{II40}$$

Here $W\{\vec{k},\vec{m},a_\mu\}$ is the Euclidean functional integral with boundary conditions fixed by choosing

---

* See footnote page II21. Read: I is the class of purely periodic Ω.

$$n_{ij} = \varepsilon_{ijk}\, m_k ; \quad n_{i4} = k_i ; \quad a_4 = \beta \, .$$

Because of the gauge choice (II39) this functional integral must include integration over the Ω belonging to the given homotopy classes as they determine the boundary conditions such as (II7). The definition of W is completely Euclidean symmetric. In the next chapter I show how to make use of this symmetry with respect to rotation over 90^o in Euclidean space.

II7. DUALITY

The Euclidean symmetry in eq. (II40) suggests to consider the following SO(4) rotation:

$$\begin{bmatrix} 0 & -1 & & \\ 1 & 0 & & \\ & & 0 & 1 \\ & & -1 & 0 \end{bmatrix} . \qquad \text{(II41)}$$

Let us introduce a notation for the first two components of a vector:

$$x_\mu = (\vec{x}, x_4) ;$$
$$\tilde{x} = (x_1, x_2) ,$$
$$\hat{x} = (x_2, x_1) . \qquad \text{(II42)}$$

We have, from eq. (II40):

$$\exp\left[-\beta F(\tilde{e}, e_3, \tilde{m}, m_3, \tilde{a}, a_3, \beta)\right] =$$

$$\frac{1}{N^2} \sum_{\tilde{k},\tilde{\ell}} \exp\left[\frac{2\pi i}{N}\left[-(\tilde{k}.\tilde{e}) + (\tilde{\ell}.\tilde{m})\right] - a_3 F(\tilde{\ell}, -e_3, \tilde{k}, -m_3, \hat{a}, \beta, a_3)\right] . \qquad \text{(II43)}$$

Notice that in this formula the transverse electric and magnetic fluxes are Fourier transformed and interchange positions. Notice also that, apart from a sign difference, there is a complete electric-magnetic symmetry in this expression, in spite of the fact that the definition of F in terms of W was not so symmetric. Eq. (II43) is an exact property of our system. No approximation was made. We refer to it as "duality".

II8. LONG-DISTANCE BEHAVIOR COMPATIBLE WITH DUALITY

Let us now assume that the theory has a mass gap. No massless particles occur. Then asymptotic behavior at large distances will be approached exponentially. Then it is exluded that

$$F(\vec{e},\vec{m},\vec{a},\beta) \to o, \text{ exponentially as } \vec{a}, \beta \to \infty ,$$

for all $\vec{e}$ and $\vec{m}$, which would clearly contradict (II43). This means that at least some of the flux configurations must get a large energy content as $\vec{a}$, $\beta \to \infty$. These flux lines apparently cannot spread out and because they were created along curves C it is practically inescapable that they get a total energy which will be proportional to their length:

$$E = \lim_{\beta\to o} F = Ca \quad . \tag{II44}$$

However, duality will never enable us to determine whether it are the electric or the magnetic flux lines that behave this way. From the requirement that W in II40 is always positive one can deduce the impossibility of a third option, namely that only exotic combinations of electric and magnetic fluxes behave as strings[4].

For further information we must make the physically quite plausible assumption of "factorizability":

$$F(\vec{e},\vec{m}) \Rightarrow F_e(\vec{e}) + F_m(\vec{m}) \quad \text{when } \vec{a}, \beta \to \infty \ . \tag{II45}$$

Suppose that we have confinement in the electric domain:

$$F_e(o,o,1) \Rightarrow \rho a_3 \tag{II46}$$

where ρ is the fundamental string constant. Then we can derive from duality the behavior of $F_m(\vec{m})$.

First we improve (II46) by applying statistical mechanics to obtain F_e for large but finite β. One obtains:

$$e^{-\beta F_e(e_1,e_2,o,\vec{a},\beta) + C(\vec{a},\beta)}$$

$$= \sum_{n_1^\pm,n_2^\pm} \frac{1}{n_1^+!n_1^-!n_2^+!n_2^-!} \gamma_1^{n_1^+ + n_1^-} \gamma_2^{n_2^+ - n_2^-} \delta_N(n_1^+ - n_1^- - e_1)\, \delta_N(n_2^+ - n_2^- - e_2) . \tag{II47}$$

Here

$$\gamma_1 = \lambda\, a_2 a_3\, e^{-\beta\rho a_1} ,$$

$$\gamma_2 = \lambda\, a_1 a_3\, e^{\beta\rho a_2} ,$$

$$\delta_N(x) = \frac{1}{N} \sum_{k=o}^{N-1} e^{2\pi ikx/N} ,$$

$$\delta_N(x) = \begin{cases} 1 & \text{if } x = o \pmod N, \\ o & \text{otherwise.} \end{cases} \qquad \text{(II48)}$$

The sum is over all nonnegative integer values of $n_i^{\pm}$ (the orientations ± are needed if $N \geqslant 3$). The γ's are Boltzmann factors associated with each string-like flux tube.

We now insert this, with (II45), into (II43) putting $e_3 = m_3 = o$. One obtains

$$e^{-\beta F_m(m_1,m_2,o,\vec{a},\beta)} = C' e^{2\sum_a \gamma'_a \cos(2m_a\pi/N)} \qquad \text{(II49)}$$

where C' is again a constant and

$$\gamma'_1 = \lambda\, a_1 \beta\, e^{-\rho a_2 a_3} ,$$

$$\gamma'_2 = \lambda\, a_2 \beta\, e^{-\rho a_1 a_3} . \qquad \text{(II50)}$$

At $\beta \to \infty$ we get

$$F_m(\tilde{m},o,\vec{a},\beta) \to E_m(\tilde{m},o,\vec{a}) = \sum_i E_i(m_i,\vec{a})$$

with

$$E_1(m_1,\vec{a}) = 2\lambda \left(1 - \cos \frac{2\pi m_1}{N}\right) a_1\, e^{-\rho a_2 a_3} . \qquad \text{(II51)}$$

and similarly for E_2 and E_3.

One reads off from eq. (II51) that there will be no magnetic confinement, because if we let the box become wider the exponential factor

$$e^{-\rho a_2 a_3}$$

causes a rapid decrease of the energy of the magnetic flux. Notice the occurrence of the string constant ρ in there.

Of course we could equally well have started from the presumption that there were magnetic confinement. One then would conclude that there would be no electric confinement, because then the electric flux would have an energy given by (II51).

II9. THE COULOMB PHASE

To see what might happen in the absence of a mass gap one could study the (first) Georgi-Glashow model[7]. Here SU(2) is "broken spontaneously" into U(1) by an isospin one Higgs field. Ordinary perturbation expansion tells us what happens in the infrared limit. There are electrically charged particles: $W^{\pm}$ (the charged vector particles). They carry two fundamental electric flux units ("quarks" with isospin $\frac{1}{2}$ would have the fundamental flux unit $q_o = \pm\frac{1}{2}e$). There are also magnetically charged particles (monopoles,[8]). They also carry two fundamental magnetic flux units:

$$g = \frac{2\pi}{q_o} = \frac{4\pi}{e} . \qquad \text{(II52)}$$

A given electric flux configuration of k flux units would have an energy

$$E = \frac{q_o^2 k^2 a_1}{2a_2 a_3} . \qquad \text{(II53)}$$

A finite β however pair creation of $W^{\pm}$ takes place, so that we should take a statistical average over various values of the flux. Flux is only rigorously defined modulo $2q_o$. We have

$$e^{-\beta F_e(1,o,o)} = \frac{\sum\limits_{k=-\infty}^{\infty} \exp\left[-\beta \frac{e^2 a_1}{2a_2 a_3} (k+\frac{1}{2})^2\right]}{\sum\limits_{k=-\infty}^{\infty} \exp\left[-\beta \frac{e^2 a_1}{2a_2 a_3} k^2\right]} . \qquad \text{(II54)}$$

Similarly, because of pair creation of magnetic monopoles

$$e^{-\beta F_m(1,o,o)} = \frac{\sum\limits_{k=-\infty}^{\infty} \exp\left[-\beta \frac{8\pi^2 a_1}{g^2 a_2 a_3} (k+\frac{1}{2})^2\right]}{\sum\limits_{k=-\infty}^{\infty} \exp\left[-\beta \frac{8\pi^2 a_1}{g^2 a_2 a_3} k^2\right]} . \qquad \text{(II55)}$$

These expressions do satisfy duality, eq. (II43). This is easily verified when one observes that

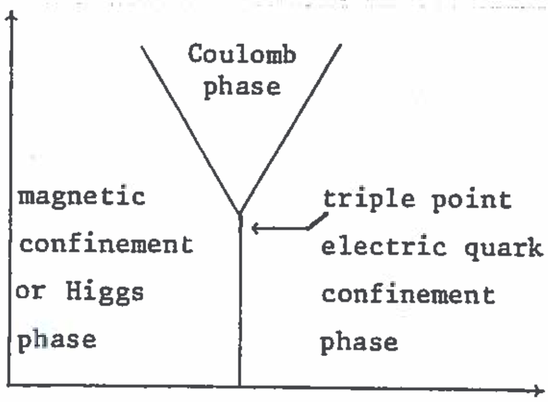
$$\sum_{k=-\infty}^{\infty} e^{-\lambda k^2} = \sqrt{\frac{\pi}{\lambda}} \sum_{k=-\infty}^{\infty} e^{-\pi^2 k^2/\lambda} ,$$

and

$$\sum_{k=-\infty}^{\infty} (-1)^k \, e^{-\lambda k^2} = \sqrt{\frac{\pi}{\lambda}} \sum_{k=-\infty}^{\infty} e^{-\pi^2 (k+\frac{1}{2})^2/\lambda} .$$

Notice now that this model realizes the dual formula in a symmetric way, contrary to the case that there is a mass gap. This dually symmetric mode will be referred to as the "Coulomb phase" or "Georgi-Glashow phase".

Suppose that Quantum Chromodynamics would be enriched with two free parameters that would not destroy the basic topological features (for instance the mass of some heavy scalar fields in the adjoint representation). Then we would have a phase diagram as in the Figure below.



Numerical calculations[9] suggest that the phase transition between the two confinement modes is a first order one. Real QCD is represented by one point in this diagram. Where will that point be? If it were in the Coulomb phase there would be long range, strongly interacting Abelian gluons contrary to experiment. In the Higgs mode quarks would have finite mass and escape easily. It could be still in the Higgs phase but very close to the border line with the confinement mode. If the phase transition were a second order one then that would imply long range correlation effects requiring light physical gluons. Again: they are not observed experimentally. If, which is more likely, the phase transition is a first order one then even close to the border line not even approximate confinement would take place: quarks would be produced copiously. There is only one possibility: we are in the confinement mode. Electric flux lines cannot spread out. Quark confinement is absolute.

REFERENCES AT LECTURE II

1. E.S. Abers and B.W. Lee, Physics Reports 9C, no 1.
   S. Coleman in: Laws of Hadronic Matter, Erice July 1973, ed. by A. Zichichi, Academic Press, New York and London.
2. H.B. Nielsen and P. Olesen, Nucl. Phys. B61 (1973) 45, ibidem B160 (1979) 380.
3. B. Zumino, in Renormalizatio and Invariance in Quantum Field Theory, ed. E.R. Caianiello, Plenum Press, New York, p. 367.
4. G. 't Hooft, Nucl. Physics B153 (1979) 141.
5. G. 't Hooft, Nucl. Physics B138 (1978) 1.
6. E. Witten, Harvard preprint (1979).
7. M. Georgi and S.L. Glashow, Phys. Rev. Lett. 28 (19732) 1494.
8. G. 't Hooft, Nucl. Physics B79 (1974) 276.
   A.M. Polyakov, JETP Lett. 20 (1974) 194.
9. M. Creutz, L. Jacobs and C. Rebbi, Phys. Rev. D20 (1979) 1915.