The Topological Mechanism for Permanent Quark Confinement in a Non-Abelian Gauge Theory

G. 't Hooft

Institute for Theoretical Physics, Princetonplein 5, P.O. Box 80006, 3508 TA Utrecht, The Netherlands

Received December 19, 1980; accepted February 3, 1981

Abstract

Permanent quark confinement occurs in a gauge theory if its vacuum (= ground state) condenses into a state that bears close resemblance to the superconductivity state of a solid. We first argue that indeed superconductivity is an absolute phenomenon, the conductivity is rigorously zero. A superconductor forces absolute confinement of magnetic charges. Then we show that a non-Abelian gauge theory may be seen as an Abelian theory enriched with Dirac magnetic monopoles. If these monopoles move collectively, as Cooper pairs do inside a superconductor, then the role of magnetism and electricity are interchanged. It is shown that then quarks are permanently confined. In SU(3) this confinement goes according to a triality rule, allowing the usual baryonic and mesonic states.

1. Introduction

The aim of these lectures is to provide an intuitive understanding of the quark confinement phenomenon in gauge theories. It is not just a trick invented by theoretical physicists to avoid confrontation with experiment: quarks, the fundamental constituents in “quantum chromodynamics (QCD)”, have never been observed as isolated particles. Permanent confinement of quarks is an absolute property of the vacuum in QCD. It is in fact closely related to another absolute law of physics: the infinite electrical conductivity of certain solids at low temperatures: “superconductivity”.

In Fig. 1 we see the contents of these lectures displayed diagrammatically. The superconductor, which, in high energy terms, is just an Abelian Higgs theory, is a medium that does not allow the presence of single magnetic charges (“magnetic monopoles”). North and south poles would attract each other in a linearly rising potential well. However, monopoles do not exist as far as we know. They would have to be introduced by imagination. We explain how P. A. M. Dirac imagines magnetic monopoles.

Although the quantum theory treats electric and magnetic charges very differently, it is reasonable to assume that they are equivalent. A “duality transformation” interchanges them. Such a dual transformation would turn a superconductor into a magnetic superconductor. It could serve as a model for the QCD vacuum. Electrically charged “quarks” will be permanently confined.

Dirac’s construction of monopoles is rather artificial. To the contrary, we explain how monopoles will occur naturally in a non-Abelian model, where the non-Abelian local symmetry is broken down by the Higgs mechanism into an electromagnetic U(1) symmetry (Georgi–Glashow model).

Those monopoles are heavy, however. If we want them to become superconducting we need “ultralight” ones. We turn back to the pure non-Abelian gauge theory without Higgs

mechanism. Now we can show that, in a certain sense, it contains monopoles that may or may not become the charge carriers of a superconductor. In general, this system may condense into three types of vacuum: (1) the ordinary superconductor or Higgs mode. Only color-magnetic charges will be confined. (2) The Coulomb mode, featuring ordinary massless photons and no superconductivity or confinement. (3) The magnetic superconductor in which quarks are permanently confined (“electric super-insulator”).

It is the details of the dynamics of the system that determine which mode is actually chosen. We do not consider that question here. Our mathematics is intuitive rather than complete. A more satisfactory mathematical categorization of the three condensation modes was discussed in the fourth lecture at the School. A lengthy account of that can be found in the Schlading lecture notes 1980 [1].

2. Maxwell’s equations

Let us recall the Maxwell equations in a vacuum, putting $c = 1$

$$\operatorname{div} \mathbf{E} = \rho \quad (1.1a)$$

$$\operatorname{curl} \mathbf{B} = \mathbf{j} + \dot{\mathbf{E}} \quad (1.1b)$$

$$\operatorname{curl} \mathbf{E} + \dot{\mathbf{B}} = 0 \quad (1.1c)$$

$$\operatorname{div} \mathbf{B} = 0 \quad (1.1d)$$

The first two of these are the inhomogeneous Maxwell equations. They allow for the presence of electric charges and currents. The two other equations tell us that there are no magnetic charges or currents. They are crucial for the gauge concept. Why and how nevertheless magnetic monopoles can be introduced will be explained later.

The components of $\mathbf{E}$ and $\mathbf{B}$ constitute six kinetic variables at each space–time point. We can reduce that to four in the following way: because $\operatorname{div} \mathbf{B} = 0$ there is a field $\mathbf{A}$ with

$$\mathbf{B} = \operatorname{curl} \mathbf{A} \quad (1.2)$$

and because $\operatorname{curl} (\mathbf{E} + \dot{\mathbf{A}}) = 0$ there is a scalar field ϕ with $\mathbf{E} + \dot{\mathbf{A}} = -\operatorname{grad} \phi$, or

$$\mathbf{E} = -\operatorname{grad} \phi - \dot{\mathbf{A}}. \quad (1.3)$$

Of course it is convenient to write these things in a relativistically covariant way:

$$x_\mu = \begin{pmatrix} \mathbf{x} \\ it \end{pmatrix} \quad j_\mu = \begin{pmatrix} \mathbf{j} \\ i\rho \end{pmatrix} \quad (1.4a)$$

Fig. 1.

$$F_{\mu\nu} = \begin{pmatrix} 0 & B_3 & -B_2 & -iE_1 \\ -B_3 & 0 & B_1 & -iE_2 \\ B_2 & -B_1 & 0 & -iE_3 \\ iE_1 & iE_2 & iE_3 & 0 \end{pmatrix} \quad (1.4b)$$

$$A_\mu = \begin{pmatrix} \mathbf{A} \\ i\phi \end{pmatrix} \quad (1.4c)$$

turning our equations into

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

$$\partial_\mu F_{\mu\nu} = -j_\nu$$

3. Abelian gauge transformations

The fields $\mathbf{A}$ and ϕ of the previous section can always be constructed, but they are not uniquely defined. Consider namely an arbitrary function $\Lambda(\mathbf{x}, t)$ and make the following change in $\mathbf{A}$ and ϕ :

$$\begin{aligned} \mathbf{A}' &= \mathbf{A} + \text{grad } \Lambda \\ \phi' &= \phi - \dot{\Lambda} \end{aligned} \quad (3.1)$$

or

$$A'_\mu = A_\mu + \partial_\mu \Lambda \quad (3.2)$$

Since $\text{curl grad} = 0$, the field $\mathbf{B}$ in eq. (1.2) does not change.

Fig. 2.

Also, $\mathbf{E}$ is not affected, according to eq. (1.3). The fields A_μ and A'_μ generate the same electro-magnetic fields $F_{\mu\nu}$. The transformation (3.1) and (3.2) is called gauge transformation. Because Λ may depend on space-time the gauge transformation is local.

In solving the classical electrodynamical equations the transition to potential fields is useful mathematically, but not crucial. However, when we try to formulate a Schrödinger equation for electrically charged particles in electro-magnetic fields then the vector potential is unavoidable. This is illustrated in Fig. 2. The shaded region M is a magnetic field orthogonal to the plane of the paper. A wave front of the electron ψ field enters from the left. The electrons are deflected. In an attempt to recover this deflection through a wave equation we assume that the magnetic field causes a refraction. However, then the wavelength at A must be different from the wavelength at B . Neither at A nor at B is a magnetic field. The only explanation of this difference is the vector potential $\mathbf{A}$ (which is different at A and B , since its curl must generate the field). $\mathbf{A}$ must occur explicitly in the Schrödinger equation.

In classical mechanics the magnetic field is introduced by substituting $\mathbf{p}$ by $\mathbf{p} + e\mathbf{A}$ in the Hamiltonian. e is the electric charge of the particle. In quantum mechanics this corresponds to the replacement

$$\partial_\mu \psi \rightarrow (\partial_\mu - ieA_\mu)\psi \stackrel{\text{def}}{=} D_\mu \psi \quad (3.3)$$

in the Schrödinger equation. Does this explicit occurrence of the vector potential not disturb gauge invariance? Gauge invariance can be recovered if the transformation (3.1) and (3.2) is associated with a transformation of ψ :

$$\psi' = e^{ie\Lambda}\psi \quad (3.4)$$

We then have

$$(D_\mu \psi)' = e^{ie\Lambda} D_\mu \psi \quad (3.5)$$

where the gradient of Λ cancels out.

This way, the Dirac equation for instance can be made totally gauge invariant. The electric current,

$$j_\mu = -ie\bar{\psi}\gamma_\mu\psi \quad (3.6)$$

is gauge-invariant because

$$\bar{\psi}' = e^{-ie\Lambda}\bar{\psi} \quad (3.7)$$

4. The Abelian superconductor

In the absence of electric charges and currents, Maxwell's equations describe the propagation of waves. These waves are quantized into photons, massless particles moving with the speed of light. The properties of photons change drastically however if

Fig. 3.

we allow them to enter a superconducting piece of material. Let us first consider the superconductor classically.

Infinite conductivity would imply that any electric field $\mathbf{E}$ would cause an "infinite" current:

$$\mathbf{j} = \infty \cdot \mathbf{E} \quad (4.1)$$

which is of course absurd. Rather, one has to postulate that no electric field can exist inside a superconductor:

$$\mathbf{E} = 0 \quad (4.2)$$

Maxwell's equation (1.1c) now tells us that

$$\dot{\mathbf{B}} = -\text{curl } \mathbf{E} = 0 \quad (4.3)$$

In other words: $\mathbf{B}$ is fixed with respect to the atoms of the superconductor. If no field was frozen in from the start, the field $\mathbf{B}$ will remain zero even if we switch on a magnet (Fig. 3). This is called the Meissner effect.

In fact, in the superconductor electric currents were induced that cause the deviation of the magnetic field lines. Now in spite of what eq. (4.1) suggests, the superconducting current is limited in strength, naturally, since there are only a finite number of electrons. This implies that if the external magnetic field becomes too strong, the superconducting material can no longer repel it completely. One finds that small regions in the superconductor make a transition to a "non-superconducting" state, such that magnetic flux lines are allowed in. Since such a transition takes energy, the magnetic regions are kept as small as possible, see fig. 4. Narrow flux tubes are formed. Their geometry is such that a compromise is found between magnetic field energy and the energy needed for the transition to the normal state. The flux tubes satisfy a quantization rule that can only be understood in terms of a microscopic, quantum-mechanical theory. Fortunately such a theory is at hand.

5. Superconductor theory, or: the Higgs mechanism

Consider the atoms of some solid material forming an ideal periodic lattice. The characteristic features of the Schrödinger equation for an electron in such an ideal lattice are well known from solid state theory: the electron moves as a free particle, but with a velocity-dependent mass. Any deviation from the purely periodic arrangement of the atoms disturbs this motion:

Fig. 4.

electrons can scatter against impurities of the lattice. In practice the lattice will be far from ideal. In particular, any sound wave causes complicated deviations. Consequently, electrons may interact with sound waves. The sound waves are quantized into so-called phonons. Phonons can be created and absorbed by electrons, much like photons in the theory of quantum-electrodynamics.

Phonons form a mixture of scalar and vector fields. Some of these cause attraction, others repulsion among electrons, in competition with the ordinary electric forces transmitted by photons.

The Bardeen-Schrieffer-Cooper theory of superconductivity [2] assumes that the attractive forces due to phonons can be strong enough to cause bound states of two electrons with opposite spin and same charge. The lowest state is a scalar particle with spin zero and charge $-2e$. This object is precisely what high-energy physicists would call a Higgs particle.

An operator that annihilates a Higgs particle can be formed from the product of two electron annihilation operators, one with spin up and one with spin down:

$$H = \psi^\dagger \psi^\downarrow = \frac{1}{2} \epsilon^{ij} \psi_i \psi_j \quad (5.1)$$

where $\psi^\dagger = \psi_1$ annihilates an electron with spin up and $\psi^\downarrow = \psi_2$ one with spin down.

H would be called "Higgs field" in particle physics, or "order parameter" in solid state physics.

Under a gauge transformation

$$\psi_i \rightarrow e^{ie\Lambda} \psi_i \quad (5.2)$$

we have

$$H \rightarrow e^{2ie\Lambda} H = e^{-iQ\Lambda} H \quad (5.3)$$

reflecting the fact that the electric charge Q of the Higgs field is $-2e$.

The H field will interact with electrons, phonons and photons in a complicated way. But to understand the basic features of a superconductor we only need to consider its self-interaction and the interaction with the electromagnetic field resulting from the electric charge of H . To simplify things we may consider a little model exhibiting these interactions. Let us consider H to be a locally defined canonical variable, satisfying canonical commutation rules (this also is a simplification of the truth) and take the Hamilton density $\mathcal{H}$ to be

$$\mathcal{H} = \frac{1}{2} \dot{H}^2 + \frac{1}{2} E^2 + \frac{1}{2} B^2 + |(\nabla + 2ei\mathbf{A})H|^2 + V(|H|) \quad (5.4)$$

In here, the number $-2e$ represents the electric charge of the H field, and $V(|H|)$ is a self-interaction. We choose V to depend only on $|H|$ so that this interaction will be automatically gauge-invariant. For the superconductor we assume V to have the shape sketched in Fig. 5. In our model just take

$$V(|H|) = \lambda(H^*H - F^2)^2 \quad (5.5)$$

Here, F denotes the value of $|H|$ where V is minimal.

If the temperature T drops below a critical value T_c then a kind of condensation takes place usually referred to as "Higgs mechanism". In the condensed phase, physics is described by performing a perturbation expansion of H around the value F :

$$H = F + h \quad (5.6)$$

expand for small h . Usually one reads in the literature that the vacuum expectation value

Fig. 5.

$$\langle H \rangle_0 = F \neq 0 \quad (5.7)$$

That would be the correct criterion for condensation in the limit $e \rightarrow 0$, where the gauge symmetry is not local but only global. But eq. (5.7) is misleading in the local gauge theory, because it is not a gauge-invariant statement. In practice it works whenever one can expand for small h and small e (keeping the product eF fixed).

Neglecting h altogether we find that the Hamiltonian density is

$$\mathcal{H} = \frac{1}{2} \dot{H}^2 + \frac{1}{2} E^2 + \frac{1}{2} B^2 + \frac{1}{2} \mu^2 A^2 + \dots \quad (5.8)$$

with

$$\mu = 2\sqrt{2}eF \quad (5.9)$$

Solving the corresponding Hamilton equations one finds the wave equations for massive vector particles. The photon, whose field is still A , received an additional term in eq. (5.8), so that it became massive. Its mass is μ . Static electrodynamic fields without external sources satisfy

$$(\nabla^2 - \mu^2) \begin{bmatrix} \mathbf{E} \\ \mathbf{B} \end{bmatrix} = 0 \quad (5.10)$$

so that they drop exponentially with distance just like the Yukawa forces in a nucleus due to exchange of massive pions.

Equation (5.10) forbids any static electric or magnetic field inside a superconductor, apart from a boundary effect: the exponential solutions allow a field to penetrate a superconductor by a penetration depth roughly equal to μ^{-1} . The conclusion of this section is that indeed the so-called Higgs mechanism corresponds to a transition towards a superconducting state.

6. The Meissner effect and flux quantization

If indeed the microscopic theory of the previous chapter explains the absence of electromagnetic fields inside a superconductor, then how should we account for the phenomena described in Section 4 when very strong magnetic fields are allowed in, as sketched in Fig. 4? This we can readily explain. We had assumed that perturbation expansion could be performed in the neighbourhood of the case $H = F$ everywhere. Instead, we could also have considered $H = F e^{i\phi}$ where ϕ is a space (-time) dependent phase factor. By (local) gauge invariance this would

Fig. 6. A possible Higgs field configuration.

be an equivalent choice, as long as ϕ is continuous and differentiable everywhere.

Not entirely equivalent is the choice depicted in Fig. 6. Here, ϕ makes a full rotation from 0 to 2π if we follow its value along the boundary of the picture [3]. The picture shows the Higgs field in a two-dimensional plane, cut somewhere through a superconducting piece of material. Let us choose these values such that the total energy is minimal, so that the remaining oscillations are entirely quantummechanical and therefore, to some extent, negligible.

Observe that away from the center the Higgs field is

$$H \rightarrow \left(\frac{-y + ix}{r} \right) F \quad (6.1)$$

If there would be no vector potential the total energy would be

$$U = \int \mathcal{H} d^2x \rightarrow \int |\nabla H|^2 \rightarrow \int \frac{2\pi r dr}{r^2} F^2 \quad (6.2)$$

which diverges logarithmically to ∞ as $r \rightarrow \infty$. But we can carefully adjust the vector potential:

$$\mathbf{A} \rightarrow \frac{1}{2er^2} \begin{pmatrix} y \\ -x \end{pmatrix} \quad (6.3)$$

such that

$$\mathcal{H} \rightarrow |(\nabla + 2ei\mathbf{A})H|^2 \rightarrow 0 \quad (6.4)$$

very rapidly (in fact exponentially).

However, it is easy to deduce from (6.3) that there would be first of all a finite magnetic flux:

$$\Phi = \oint \mathbf{A} d\mathbf{x} = \frac{\pi}{e} \quad (6.5)$$

implying the existence of a magnetic field with a nonvanishing amount of energy. Secondly, we should require that both H and $\mathbf{A}$ are continuous everywhere, the origin included, so that eqs. (6.1) and (6.3) cannot hold everywhere. Somewhere close to the origin, H must have a zero point. At that point $V(|H|)$ is larger than in the vacuum, so that also V contributes to the energy.

In fact what we have found is a new, stable field configuration representing a magnetic flux tube, of which our two-dimensional sheet shows a cross section. The flux tube carries a non vanishing amount of energy per unit of length. It has a definite width, in the order of μ^{-1} and the total amount of magnetic flux is fixed to be π/e . Of course we could envisage several of these flux tubes touching each other, a situation that would

correspond to a picture in which the Higgs field makes multiple rotations when we follow it at the boundary. Thus we find that magnetic flux inside a superconductor does not necessarily vanish but is quantized:

$$\Phi = \frac{n\pi}{e}, \quad n = \dots, -2, -1, 0, 1, 2, \dots \quad (6.6)$$

Since these flux tubes carry a significant amount of energy per unit of length, they only appear if a superconductor is immersed in a magnetic field that exceeds a certain critical value. Obviously these flux tubes are locally stable, that is, they cannot break because, among other things, magnetic flux is conserved. Their only end points may be at the surface of the piece of superconducting material, where they merge into a normal magnetic field.

7. Magnetic monopoles and permanent magnetic confinement

But imagine that a little green man from outer space handed to us a magnetic monopole, a particle with magnetic north charge without a south pole attached, and another one gave us an anti-monopole being a pure magnetic south charge. If the total magnetic charge of each monopole would correspond to magnetic flux $\pm \pi/e$, then if these monopoles would be immersed into the superconductor, they would form end points of Meissner flux tubes. Since those flux tubes carry a definite amount of energy per unit length the total energy of the system would be at least proportional to the separation length of the monopoles. In other words: There would be a linearly rising potential pulling the monopole-anti-monopole pair together regardless how far they are separated: magnetic monopoles inside a superconductor are permanently confined. Indeed, this permanent confinement law looks suspiciously like permanent quark confinement the way it is believed to occur in quantum chromodynamics [4]. But we first must answer two questions. One: how can we introduce magnetic monopoles in a theory built upon the use of vector potentials, whose very existence was due to the absence of magnetic charges, and two: why would the magnetic charge of such monopoles be precisely (or perhaps a multiple of) π/e ? If it would be a fraction of that charge then it would require a Meissner vortex with fractional magnetic flux, which we have not been able to construct. We will answer both questions simultaneously in the next chapter.

8. Dirac monopoles in the Abelian gauge theory

In spite of what we said in Chapter 3, Paul Dirac was able to introduce a theory for a purely magnetically charged particle which is not in contradiction with quantum mechanics [5]. His argument amounts to the following. We would like to write down a wave equation for electrically charged particles even if the magnetic field has a static source. We do this by mimicking that magnetic charge in a Gedanken experiment. A very long and thin, densely and evenly wound coil stretches from infinity to the point of interest, say the origin of a coordinate system (Fig. 7). An electric current through the coil produces the same field configuration as a magnetic monopole, except that the total magnetic flux is conserved due to a strong magnetic field *inside* the coil. The question is now: can we arrange the coil such that in the limit of the coil becoming infinitely thin it becomes invisible to the electrons?

Let the desired magnetic field be

Fig. 7. Coil producing the field of a monopole.

$$\mathbf{B} = \frac{g}{4\pi r^3} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (8.1)$$

Here the unit of magnetic charge g is chosen such that the force law for two magnetic charges is just as Coulomb's law. If the coil runs in the positive z direction then the vector potential may be chosen to be

$$\mathbf{A} = \frac{g}{4\pi} \frac{z+r}{\rho^2 r} \begin{pmatrix} y \\ -x \\ 0 \end{pmatrix} \quad (8.2)$$

where $r^2 = \rho^2 + z^2 = x^2 + y^2 + z^2$.

Clearly this expression shows a singularity at $\rho \rightarrow 0, z \rightarrow +r$ but no singularity at $\rho \rightarrow 0, z \rightarrow -r$. At $z \simeq r$ we have:

$$\mathbf{A} \rightarrow \frac{g}{2\pi\rho^2} \begin{pmatrix} y \\ -x \\ 0 \end{pmatrix} \quad \rho^2 = x^2 + y^2 \quad (8.3)$$

Let us now perform the following singular gauge transformation:

$$\begin{aligned} \psi &\rightarrow e^{i\theta n} \psi \\ A_\mu &\rightarrow A_\mu + \frac{n}{e} \partial_\mu \theta \end{aligned} \quad (8.4)$$

where θ is the angle in the x - y plane:

$$x + iy \equiv \rho e^{i\theta} \quad (8.5)$$

In other words, the gauge function Λ is

$$\Lambda = \frac{\theta n}{e} \quad (8.6)$$

What makes this choice remarkable is that Λ is not single valued: it jumps by $2\pi n/e$ when we make a full turn about the z -axis. However since ψ and A_μ all remain single valued the Schrödinger equation before and after this gauge transformation

remains the same. Any non-integer multiple of θ/e would be unacceptable for Λ .

The new Schrödinger equation has

$$\mathbf{A} \rightarrow \left(\frac{g}{2\pi} - \frac{n}{e} \right) \frac{1}{\rho^2} \begin{pmatrix} y \\ -x \\ 0 \end{pmatrix}. \quad (8.7)$$

If $g/2\pi = n/e$ for some integer n , or

$$ge = 2\pi n \quad (8.8)$$

then we can remove the singularity of $\mathbf{A}$ by a gauge transformation. We do get it back somewhere else such as for instance the negative z -axis (if eq. (8.4) is continued everywhere), but we did show that no observer can claim he saw the coil, because it can be transformed away from him. This way, the monopole becomes rotationally symmetric and singularity free (except of course at the origin itself) and therefore physically acceptable. Note that eq. (8.8) implies that the magnetic charge quantum is 68.5 times the electric charge quantum and exactly twice the superconductor flux quantum. The latter factor two arises from the fact that the Higgs field in the superconductor carries two units of electric charge.

9. Dual symmetry

Let us assume for a moment that Dirac's monopoles really exist. This would imply that those Maxwell equations that originally were homogeneous, eqs. (1.1c) and (1.1d) now obtain an inhomogeneous right hand side: the magnetic currents and charges. Maxwell's equations are now much more symmetric than they used to be. However, in quantizing the system we have put electric and magnetic charges on a very different footing: electrons were light, their charge e is small, so that perturbation expansion in quantum electrodynamics makes sense, and complete second quantization can be carried out: electrons and positrons are copiously created and annihilated. On the other hand, the magnetic charges g are large (they have to be because $ge/2\pi$ must be a nonvanishing integer). Therefore monopoles are likely to be heavy. They have a virtual infinite coil ("Dirac string") attached to them. Because of the large value of g , second quantization is extremely hard (without any further knowledge of the inner structure of the monopole). But formally, yes, the starting point of our theory, the Maxwell equations, were symmetric under the rotation

$$\mathbf{E} \rightarrow \mathbf{E} \cos \theta + \mathbf{B} \sin \theta \quad (9.1)$$

$$\mathbf{B} \rightarrow -\mathbf{E} \sin \theta + \mathbf{B} \cos \theta$$

or in particular the special case:

$$\mathbf{E} \rightarrow \mathbf{B} \quad (9.2)$$

$$\mathbf{B} \rightarrow -\mathbf{E}$$

Because the product of e and g is fixed we cannot have a small e and a small g expansion simultaneously. This is why any field theory of monopoles is not manifestly symmetric under eq. (9.2). We assume however, that any system of electric and magnetic charges may in principle be handled in any of the two dually opposite frameworks,

10. Permanent electric confinement

Thus, it should be possible to consider "dual" transformations in the theory. Suppose we considered ordinary quantum-

electrodynamics, enriched with Dirac magnetic monopoles. Suppose we choose a smaller and smaller mass value for these monopoles, and reduce their magnetic charge by increasing the electric charge unit. From some point on it would be more advisable to describe this system in the “dual formulation”. We ascribe elementary fields to the monopoles and treat the electrically charged particles classically, with strings attached. In such a formalism it is easy to imagine that the Higgs mechanism takes place, with respect to the fields of the *magnetically* charged particles. These fields get a non-vanishing vacuum expectation value so that a “superconductor” is obtained, but this “superconductor” only superconducts *magnetic* charges. As in ordinary superconductors, the Meissner effect takes place, but this time we get *electric* flux tubes, with quantized values for the electric flux. Any electrically charged particle will be at one of the end points of such a flux tube. And so, if this happens, then all electrically charged particles will be permanently confined. The confining potential rises linearly.

Would this be a satisfactory theory for quark confinement? In the original formalism the magnetic monopoles could not properly be quantized; after the “dual transformation”, which we have not been able to formulate precisely in terms of all particles and fields involved, the “electrons” (quarks) were not properly quantized. In neither formalism do we have an exactly defined theory. So although we have a reasonable understanding of a possible mechanism, the dynamics of the theory must still be cast into an acceptable set of equations. It is for this reason that we now turn our attention to the non-Abelian gauge theories. We will show that they contain monopoles without strings attached.

11. Non-Abelian gauge theory

The gauge transformations (3.1), (3.2) and (3.4) were Abelian gauge transformations. That means that any two of such transformations can be carried out in arbitrary order to give the same net transformation: they commute.

Non-commuting gauge transformations are possible for instance if we have fields $q(x)$ with several components (“colours”)

$$\begin{pmatrix} q'_1 \\ q'_2 \\ q'_3 \end{pmatrix} = \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \quad (11.1)$$

or in short hand notation:

$$q' = \Omega q \quad (11.2)$$

We will often choose q to have just 3 components and Ω to be a 3×3 unitary matrix with determinant one. Such transformations form a group called SU(3).

Again, the derivative of this field can be turned into a gauge-covariant derivative at the cost of the introduction of a vector-potential field [6]:

$$D_\mu q = \partial_\mu q + ig \mathbf{A}_\mu q, \quad (11.3)$$

where $\mathbf{A}_\mu$ is a matrix of 3×3 four-vectors. It transforms as

$$A'_\mu = \Omega A_\mu \Omega^{-1} + \frac{1}{ig} \Omega \partial_\mu \Omega^{-1} \quad (11.4)$$

From these starting points one can now construct a non-Abelian

version of the Maxwell equations: the by now well known Yang–Mills equations [6]. The point I want to make in these lectures is that the non-Abelian gauge theories contain field configurations that should be considered as magnetic monopoles, without the need of a Dirac string. To establish that we first make a side-track towards the so-called Georgi–Glashow model (see Fig. 1).

12. The non-Abelian Higgs mechanism

Having a non-Abelian version of electromagnetism we can imagine a non-Abelian version of a superconductor. But this time there are several distinct possibilities. Let us illustrate this for the simplest non-Abelian gauge model: the one based on the gauge group SU(2). We have two-component fields and consider rotations given by 2×2 complex unitarity matrices with determinant one. Again it is assumed that at some stage a spinless particle emerges, which may either be a bound state of more elementary constituents, or a fundamental particle. It must be a non-trivial representation under SU(2), i.e., it must have at least two components that are transformed into each other by the SU(2) matrices. The self-interactions may now again be such that the corresponding fields develop a vacuum expectation value:

$$\langle H \rangle = F \hat{H}_0 \quad (12.1)$$

where F is the length and $\hat{H}_0$ is a unit vector.

Distinct possibilities arise when one asks whether there are any non trivial gauge transformations that leave $\hat{H}_0$ invariant. Those transformations form a subgroup of the gauge group (which was here SU(2)). If H has just two complex components then that subgroup only contains the identity element. All Yang–Mills photons then become massive and the vacuum is then a complete Yang–Mills superconductor. If H was some bound state of gauge vector fields, then it would have integer isospin, and that subgroup contains at least the *center* of the gauge group. The center of a group is the largest Abelian subgroup of elements that commute with all other elements of the gauge group. The center of SU(2) contains $\pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

But we will now concentrate on the possibility that the subgroup that leaves $\hat{H}_0$ invariant is a continuous, Abelian group. At least one Yang–Mills photon will remain massless. It will produce long range fields not affected by any form of superconductivity.

An example of a model with this feature is the so-called O(3)-Georgi–Glashow model.

13. The Georgi–Glashow model

The model [7] is based on SU(2) but since we will only see states with integer isospin we can use the equivalent group O(3). There are fermions which come in triplets

$$\begin{pmatrix} E^+ \\ \nu_e \\ e^- \end{pmatrix} \quad (13.1)$$

where E^+ is a newly postulated particle, and also the gauge vector fields from triplets:

$$\begin{pmatrix} W^+ \\ \gamma \\ W^- \end{pmatrix} \quad (13.2)$$

Here γ is the ordinary photon. A triplet can be considered as three real fields, for instance

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \pm iW_2) \quad \gamma = W_3 \quad (13.3)$$

Finally we have a triplet Higgs field H , whose vacuum values become

$$\begin{pmatrix} H^+ \\ H^0 \\ H^- \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ F \\ 0 \end{pmatrix}$$

or equivalently,

$$\begin{pmatrix} H_1 \\ H_2 \\ H_3 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ F \end{pmatrix} \quad (13.4)$$

Clearly, all $O(3)$ rotations about the 3-axis leave eq. (13.4) invariant, but any charged particle field rotates just as in electromagnetism. This is why W_3 is a photon and stays massless. In short: our model contains the usual Maxwell fields and in addition a bunch of heavy particles, many of which are electrically charged. Although formally there are also non-Abelian varieties of Maxwell fields (Yang–Mills fields), they are short-ranged due to the Higgs mechanism.

There is however, something else in the model: there are “natural” magnetic monopoles. This we explain in the next chapter.

14. Magnetic monopoles in the Georgi–Glashow model

In our Abelian theory any magnetic monopole necessarily had a Dirac string attached to it. We now claim that this is no longer so if the $U(1)$ gauge group of electromagnetism is imbedded in a compact non-Abelian group such as the $O(3)$ group in the Georgi–Glashow model [8]. One can formulate the problem in the following way: how can we perform a gauge transformation that is regular in space except on half of the z -axis (it might be singular at the origin). On the singular part of the z -axis the rotation should become of the form (8.4). Such a gauge rotation would produce or remove a Dirac string, and so it would enable us to create field configurations which are regular everywhere but still contain a magnetic charge (in the $U(1)$ sense) at the origin. That it is possible to perform such rotations can be demonstrated by holding a glass of water in one hand. It is easy to rotate the glass about a vertical axis for 720° with respect to oneself. After 360° the arm holding the glass is twisted. But one can undo the twist by rotating 360° in the *same* direction. The reason is that although the glass is restricted to undergo $U(1)$ rotations only, the arm is allowed to undergo any $O(3)$ rotation.

Thus, a Dirac string corresponding to a 720° gauge rotation can be produced out of nowhere. The Higgs field H , which was $(0, 0, H_0)$ before the rotation, now ends up with a node. It gets a “hedgehog” configuration:

$$H \rightarrow \frac{H_0}{r} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (14.1)$$

Because of the double rotation the obtained monopole satisfies

$$eg = 4\pi \quad (14.2)$$

This happens to agree with a restriction predicted by Schwinger, namely that Dirac’s integer quantum had to be even. However, nobody would forbid the occurrence of particles with electric charge $q = \pm \frac{1}{2}e$ in the model; they would be isospin doublets:

$$\begin{pmatrix} \psi^{\frac{1}{2}+} \\ \psi^{\frac{1}{2}-} \end{pmatrix} \quad (14.3)$$

They do saturate Dirac’s rule:

$$qg = 2\pi \quad (14.4)$$

Once the field configurations relevant for pure magnetic charges are understood in the Georgi–Glashow model, one can calculate the corresponding mass by minimizing the total energy of such configurations, keeping the fields far from the center fixed (as a boundary condition). One readily finds that the monopole mass satisfies

$$M_{\text{mon}} = C \frac{M_w}{\alpha} \cong 137 CM_w \quad (14.5)$$

where C is a constant close to one and M_w is the mass of the heavy charged Yang–Mills photons. All other properties of the monopoles can in principle be computed by perturbation expansions in e alone.

So now we have a quantitative theory for magnetic monopoles. We have a qualitative theory about what they can do if they move collectively, by considering the system in the dually opposite framework. The problem will still be that in order for monopoles to produce a magnetic Higgs mechanism they must become very light, formally even

$$M_{\text{mon}}^2 < 0. \quad (14.6)$$

This means, considering eq. (14.5) that either $\alpha \rightarrow \infty$ or $M_w^2 \rightarrow 0$ or both. The fact that M_w^2 must go through zero suggests that the original Higgs scalar field must be removed. Its condensation must be made undone. In short: we have to return to the pure gauge theory without Higgs mechanism.

15. The Abelian projection in $SU(3)$

In this section we return to $SU(3)$ and in the same time discard the Higgs mechanism. In order to be able to discuss electric and magnetic charges it is useful to consider an Abelian subgroup, by preference the largest Abelian subgroup or Cartan subgroup [9]. Here it is $U(1) \times U(1)$, the group of diagonal unitary matrices with determinant one:

$$\Omega = \begin{pmatrix} \exp\left(\frac{i\Lambda_1}{2} + \frac{i\Lambda_2}{3}\right) & 0 & 0 \\ 0 & \exp\left(\frac{-i\Lambda_1}{2} + \frac{i\Lambda_2}{3}\right) & 0 \\ 0 & 0 & \exp\left(\frac{-2i\Lambda_2}{3}\right) \end{pmatrix} \quad (15.1)$$

Let us call this the gauge group of electromagnetism-squared. There are two corresponding neutral photons. The off-diagonal components of the vector-potential matrix correspond to six charged photons. The charges with respect to the two electromagnetic forces can be plotted in a two dimensional diagram (Fig. 8a). This way of representing things could be called an

Fig. 8. (a) ○ photon electric charges, ● quark electric charges, × anti-quark electric charges. (b) magnetic charges: ● regular monopoles, ○ singular monopoles, × singular monopoles of opposite signature.

“Abelian projection”. It must be noted that at this stage the distinction between neutral photons and charged photons is quite artificial. Any gauge transformation that does not belong to the class (15.1) transforms one into the other. So our way of attaching labels to the various particles is not gauge-invariant. It is convenient however to disentangle them this way in order to have a better qualitative understanding of what will happen.

16. Color-electric charges

At a later stage one could introduce extra fields in addition to the pure gauge theory, in particular fermion fields transforming as triplets under SU(3). They are called “quarks”. Their electric charges are

$$\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} \pm \frac{1}{2} \\ 1/3 \end{pmatrix} g \quad \text{or} \quad \begin{pmatrix} 0 \\ -2/3 \end{pmatrix} g \quad (16.1)$$

for which we used an arbitrary charge unit g . The Yang–Mills (“gluons”) are either completely neutral (in this sense) or have charges

$$\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} \pm 1 \\ 0 \end{pmatrix} g \quad \text{or} \quad \begin{pmatrix} \pm 1/2 \\ \pm 1 \end{pmatrix} g \quad (16.2)$$

Notice that the gluonic charges may be obtained by putting three quarks together, or a quark and an antiquark.

17. Color-magnetic charges

Just like the Georgi–Glashow model, this system will contain naturally magnetic monopoles. They are field configurations that are either connected to the appropriate Dirac strings or, *ad libitum*, the strings may be removed by gauge rotations outside the class (15.1). We can apply a rule of thumb that we already saw in the Georgi–Glashow case: take the smallest or most general representations of the group; find out what the smallest allowable set of U(1) charges is, and then apply Dirac’s rule, which will be saturated. The fact that it will be saturated follows from group representation theory. We now have two charges (Q_1, Q_2) , so there will be also two magnetic charges (g_1, g_2) . Dirac’s condition is

$$Q_1 g_1 + Q_2 g_2 = 2\pi n \quad (17.1)$$

If we take for (Q_1, Q_2) the various quark charge combinations we get a lattice of allowed (g_1, g_2) values:

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3/2 \end{pmatrix} n_1 + \begin{pmatrix} 1 \\ -3/2 \end{pmatrix} n_2 \quad (17.2)$$

with n_1 and n_2 integer. See Fig. 8(b). They will occur regularly in the system. Since no Higgs mechanism takes place, their field configuration will require little or no energy.

However, with respect to the pure gluon theory (without quarks) one could imagine more values for the magnetic charges. Dirac’s original theory would allow for monopoles that satisfy eq. (17.1) with for (Q_1, Q_2) only the gluonic charges substituted. Since the gluonic charge lattice is only a sublattice of the complete lattice, this will be a smaller restriction than previously. Therefore one gets a more general solution:

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix} n'_1 + \begin{pmatrix} 1 \\ -1/2 \end{pmatrix} n'_2 \quad (17.3)$$

with $n'_{1,2}$ integer. They are open circles and crosses in Fig. 8(b). However, there are strings attached to these monopoles which are impossible to remove with any SU(3) gauge rotation. They are singular monopoles that do not occur naturally in the theory. They would even be incompatible with the existence of quarks.*

18. Phases of the vacuum

The bigger monopoles of type (17.2) are present, the smaller ones, (17.3) absent. Classically one can write down field configurations with the bigger magnetic charges and arbitrarily low energy (remember that the classical theory is scale invariant, so there will be very low mass monopoles becoming very large in size). But, quantum mechanics is likely to discurb their masslessness. It can go either way: one, monopoles simply become heavy, a phenomenon typical for ordinary Higgs theories. Since at the classical level no Higgs field was introduced, we would say that in this case a dynamically generated Higgs field does the job, a situation very similar to the real superconductor, where the Higgs is also a dynamically generated bound state of two electrons (“Cooper pair”). Besides this ordinary dynamical Higgs mode there is of course the possibility that no Higgs phenomenon occurs at all in the Abelian sector. The old Georgi–Glashow model is an example of that: the non-Abelian group SO(3) is broken down spontaneously into its Abelian subgroup U(1) but not further. Magnetic and electric charges coexist, respecting peacefully Dirac’s quantization condition. This second mode could also be called “Coulomb mode” because of the surviving long-range Coulomb forces.

We are dealing with a system with electric and magnetic charges. Instead of the direct picture we can consider the Higgs mechanism after applying the “dual transformation”, attributing fields to all magnetic charges and leave the electric charges as classical spectators. We then get scenario number three: the (bigger) monopoles get a formally negative mass squared; they move collectively due to the quantum corrections, and magnetic superconductivity develops. The flux quantum is the one corresponding to the superconducting charge carriers. Since they were assumed to be the bigger monopoles (17.2), the flux tubes carry the quantum number of quarks, not gluons.

A neat picture of quark confinement has emerged. Every-

* A model in which SO(8) breaks down spontaneously into SU(3) could also allow for such monopoles. Quarks would then be forbidden.

Fig. 9.

thing was formulated in terms of the Abelian charges of the Abelian subgroups of color SU(3) (Cartan subgroup). All that is needed is that the ubiquitous gas of (bigger) magnetic monopoles with magnetic charges (17.2) becomes superconducting. Since the temperature of the vacuum may be considered to be zero this may not be such a drastic assumption.

The flux tubes are characterized by one of the charge combinations of the quarks, (16.1). Clearly they may have a quark at the end and an antiquark at the beginning, forming mesons (Fig. 9a), or they are attached to a gluon with charges (16.2) in a construction of the form given in Fig. 9(b). This is a baryon. We stress that the emerging picture of a baryon having necessarily a gluon at its center is presumably an artifact of our approach. There could be a degree of freedom in a baryon corresponding to the location where three strings meet, but calling that a gluon would be stretching the picture a bit too far.

Fig. 10.

Note that the flux tube could also be an endless one (Fig. 9c): the "glueball". It is a little bit of a mystery why the Particle Tables do not show clearly several resonances corresponding to such states. Perhaps they are very unstable against quark-pair formation. Still, we did not consider the dynamics of the required quantum forces that enforce the phase transition towards the color magnetic superconductor. This is attempted by other authors. If we leave out this question of dynamics then we must envisage the other two possibilities.

Imagine that quantum chromodynamics would be a theory with two important free parameters. If we vary these parameters the system might show phase transitions from one mode into another (Fig. 10), perhaps with a triple point. Some numerical calculations suggest that the phase transition between Higgs and Confinement mode is a first order one [10].

The real world of course has definite values for those parameters, no matter what they represent. It is therefore to be indicated by a particular point in Fig. 10. These days we think we have numerous indications that the point is in the area labelled: Confinement mode. This implies, according to our theory, that quarks are permanently confined, bound together by narrow, infinitely stretchable flux tube bonds. These bonds provide an infinitely rising potential energy as a function of interquark distance. As soon as the quarks are further apart than one Fermi or so, this potential becomes a linear function of the separation distance.

References

1. 't Hooft, G., Confinement and Topology in Non-Abelian Gauge Theories, Utrecht preprint, to be published in the Proceedings of the XIX Internationale Universitätswochen für Kernphysik, Schladming, Austria, Feb. 20–March 1 (1980).
2. Bardeen, J., Cooper, L. N., and Schrieffer, J. R., Phys. Rev. **108**, 1175 (1957).
3. Rickayzen, G., Theory of Superconductivity, Interscience, John Wiley & Sons, New York, London, Sydney (1965); Williams, J. E. C., Superconductivity and its Applications, Pion Limited, London.
4. Nielsen, H. B. and Olesen, P., Nucl. Phys. **B61**, 45 (1973); Zumino, B., in Renormalization and Invariance in Quantum Field Theory, Plenum Press, (Edited by E. R. Caianiello), p. 367.
5. Dirac, P. A. M., Proc. Roy. Soc. **A133**, 60 (1931); Phys. Rev. **74**, 817 (1948).
6. Yang, C. N. and Mills, R. L., Phys. Rev. **96**, 191 (1954).
7. Georgi, H. and Glashow, S. L., Phys. Rev. Letters **28**, 1494 (1972).
8. 't Hooft, G., Nucl. Phys. **B79**, 276 (1974); Polyakov, A. M., JETP Lett. **20**, 194. (1974).
9. For a more complete discussion, see Englert, F. and Windey, P., Phys. Rev. **D14**, 2728 (1976); Goddard, P., Nuyts, J. and Olive, D., Nucl. Phys. **B125**, 1 (1977).
10. Creutz, M., Jacobs, L. and Rebbi, C., Phys. Rev. **D20**, 1915 (1979).