NON-ABELIAN GAUGE THEORIES AND QUARK CONFINEMENT

G. 't Hooft
Institute for Theoretical Physics
University of Utrecht, The Netherlands

ABSTRACT

After some philosophical remarks concerning the development of theoretical physics in this century a short introduction to non-Abelian gauge theory is given. If we concentrate on the case where the local gauge group is SU(N) we find that gauge-invariant definitions can be given of the concepts of electric and magnetic flux. Both are quantized and conserved modulo N. Either the electric or the magnetic flux must be squeezed in narrow flux tubes called vortices or strings. If the vortices are electric then quarks are confined into grouplets of multiples of N quarks. A phase transition, separates the "quark confinement mode" from the "Higgs mode", the latter being easier accessible by the use of perturbation expansion.

INTRODUCTION

It is a great honor for me to speak here at the celebration of the anniversaries of a great physicist and a beautiful equation. I choose as my subject a tantalizing problem in present-day quantum field theory and will discuss some ideas that bring us close to its resolution. How do these ideas connect with the fundamental developments in physics early this century?

Let me summarize what, in my opinion, were the most basic discoveries made in the physical sciences. In previous centuries it was already discovered that progress could be made by performing experiments, and using objectivity rather than prejudice to determine what the laws of physics are. Thus the laws of classical mechanics and chemistry were settled and atomic physics was well on its way. The fathers of this idea were Gallileo, Newton, Huijgens, etc.

The most striking event in the beginning of this century was the development of both quantum mechanics and the theories of special and general relativity. Perhaps their near coincidence was not an accident. Although these theories look very different they are all linked to one basic new discovery: the dogma concerning the role of the observer in an experiment. In interpreting the experiment it may be crucial to realize that the observer is part of the physical system he is dealing with. Thus physical laws govern the functioning of the clocks he uses, and the photons he emits to look at a physical system may disturb the same system. It may not be possible to use a preconceived universal reference frame. So we put Einstein, Planck, Heisenberg, ... under just one heading.

Finally we have the discovery of mathematics and logics, which is being done over and over again at higher levels of sophistication and abstraction. To discover a physical law one must start with a large set of possibilities and then try to reconcile several physical requirements which might seem contradicting at first sight. The

ISSN: 0094-243X/78/001/$1.50

Dirac equation is an excellent example of an achievement made along these lines.

No truly great discovery has been made since. We still make use of only these three philosophies when dealing with the problems of our times. For example we have been facing the problem how to construct a working field theory for vector-like forces of various kinds as they seemed to govern weak, electromagnetic and even strong interactions. The answer turned out to be "non-Abelian gauge theory", originated by Yang, Mills[1] and Shaw[2] who were obviously inspired by Einstein's general relativity theory. It was much later however that the necessity of this gauge principle was shown to follow from physical requirements such as renormalizability and unitarity.

I now come to my real subject: the elusive quarks. We are trying to make a theory for particles that can never be isolated completely but nevertheless seem to be fundamental building blocks of matter. Naively, we can here apply what we learned in lesson # 2, concerning the role of the observer. We should take his readings litterally ("I do not observe any free quarks") and refrain from the use of a subjective, universal reference frame (the concept of particles making up for all matter). Conclusion: quarks are silly, and attempts to make a quark theory are as futile as those for the ether theory etc.

I argue that the above reasoning is shortsighted and wrong for various reasons. First of all "observation" is confused with "isolation as a free particle". Quarks have been observed numbers of times but always in the direct vicinity of other quarks. Secondly, the idea of permanently confined but otherwise normal particles[3] should be abandoned only if it leads to real contradictions or if a better alternative theory exists. For instance one could try to introduce quarks as objects with wrong statistics, or magnetic monopoles, or simply avoid any notion of quarks altogether. And then we should apply lesson # 3: do the mathematics and consider the physical requirements. If "asymptotic freedom"[4] is taken as one of the requirements one naturely selects out the truly confined "colored" but otherwise ordinary fermionic quarks as the only acceptable theory.

Selected by elimination. But does the theory work? My claim is yes. Confinement, by infinitely rising linear potentials, can be a property of the vacuum state precisely as basic as "superconductivity" of a solid at low temperatures. It is a new phase, a new appearance, for a system of fields, separated from another phase ("Higgs mode") by a phase transition at a critical value of certain parameters. Technical details will have been published[5] by the time these notes appear in print.

NON-ABELIAN GAUGE THEORY

The field equations in quantum-electrodynamics are invariant under an Abelian group of gauge transformations:

$$\psi \to e^{i\Lambda(x)}\psi \quad ,$$
$$A_\mu \to A_\mu - \frac{1}{e}\partial_\mu \Lambda \quad . \tag{1}$$

The gauge group is U(1) and since Λ is space-time dependent we call this a local invariance.

In the non-Abelian extension of this theory the fermion field is replaced by a multiplet of fermion (or scalar) fields and one demands invariance under a larger (non-Abelian) group of transformations[1]:

$$\psi \to \Omega(x)\psi \ . \tag{2}$$

The vector potential field then must be replaced by a multiplet of vector fields forming a Hermitean matrix. They must transform as

$$A_\mu \to \Omega(x)\, A_\mu \Omega^{-1}(x) + \frac{1}{gi}\,\Omega(x)\,\partial_\mu \Omega^{-1}(x) \ . \tag{3}$$

The generalization of the electromagnetic fields $F_{\mu\nu}$ is

$$G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu] \ . \tag{4}$$

The inhomogeneous Maxwell equations are

$$D_\mu G_{\mu\nu} \equiv \partial_\mu G_{\mu\nu} + ig[A_\mu, G_{\mu\nu}] = j_\nu \tag{5}$$

where j_ν is the current of the fermion and/or scalar fields.

NON-ABELIAN ELECTRIC FLUX

The extra commutators in eqs. (4) and (5) are special features of the non-Abelian theory. They give more non-linear terms in the field equations. Replacing the commutator in (5) to the right hand side we notice that there is an extra current due to the gauge field themselves, i.e., the gauge quanta are charged, contrary to the photons in the Abelian case. This implies that, even without fermions or scalar particles the gauge-electric field is not conserved in the usual sense. Is there any way in which we can restore the concept of a conserved electric flux in a pure gauge theory without additional fermions or scalars?

The answer to that question depends on the gauge group chosen. We will consider SU(N) where N is as yet a free integer $\geqslant 2$. (For the strong interaction theory with quarks we will have N = 3.) These groups have an invariant Abelian subgroup Z(N), whose elements are, if written as N×N matrices,

$$\Omega_n = e^{2\pi i n/N}\, I \tag{6}$$

where I is the identity matrix, and n is an integer, $o \leqslant n < N$. Obviously, the gauge quanta generated by the field A_μ are invariant under Z(N) transformations (see eq. (3)). They therefore have no Z(N) "charge". We could introduce spectator fermions ψ, for instance as elementary N-fold multiplets, which do transform under Z(N). They would carry a special gauge field configuration around them that can never be neutralized by gauge quanta. This allows one to define the concept of a conserved electric flux.

However this flux differs from ordinary electric flux in two important ways. Firstly, the flux is quantized, because the smallest unit is the amount emitted by just one N-multiplet. Secondly, if we put N of these smallest units together then we obtain a source which is again singlet under Z(N), so that its effect can be neutralized by the gauge quanta. Consequently, the conservation is only modulo N. If Φ is the integer that counts this electric flux then a more convenient symbol is

$$B = e^{2\pi i\Phi/N} \tag{7}$$

which is then multiplicatively conserved.

Now we will define[5] an operator B(C) for each closed curve C that measures (multiplicatively) the amount of Z(N) electric flux that goes through the curve C. Its eigenvalues are $e^{2\pi i n/N}$. At any point x not on C our operator B(C) will be just a gauge transformation given by some $\Omega(x)$. Since we only need to tell how B(C) acts on the gauge field we need only specify Ω up to elements of Z(N). We now choose Ω such that if x is followed along a contour C' that winds through C exactly once, labeled by a parameter $o \leqslant \theta < 2\pi$, then

$$\Omega(\theta=2\pi) = e^{2\pi i/N}\,\Omega(\theta=o) \; . \tag{8}$$

Not only does this operator have the desired multiplicative properties and $[B(C)]^N = I$ (since physical states are invariant under single-valued gauge transformations*), but one can also verify that in the presence of an elementary "spectator" N-fold multiplet particle the flux changes by one unit if the particle goes through the curve C.

NON-ABELIAN MAGNETIC FLUX

We have a similar difficulty when we try to generalize the concept of magnetic flux to the non-Abelian case. The Abelian formula was simple: the flux through a closed curve C is

$$\Phi = \oint_C A_i dx^i \; . \tag{9}$$

But in the non-Abelian case this is not gauge-invariant, therefore not an observable quantity. A gauge invariant operator, defined in terms of the vector potential field on a closed curve, does exist:

$$A(C) = \mathrm{Tr}\, P \exp[ig \oint_C A_i dx^i] \; . \tag{10}$$

Here P stands for path ordering of the exponent the integral. If we write

*The necessaty of an infinitesimal smearing of the infinities arising on the curve C might cause complications here which we will ignore.

$$A(C) = N\, e^{2\pi i\Phi/N}$$

then Φ is only defined modulo N, just as in the electric case. The eigenvalues of A(C) are not so restricted as in the electric case, except when the curve C is in a vacuum: on C we have then

$$A_\mu = \frac{1}{gi}\, \Omega(x)\, \partial_\mu \Omega^{-1}(x).$$

If we do not insist on having a vacuum at the interior of C then this Ω may be multivalued, so that we obtain

$$A(C) = N\, e^{2\pi i n/N} , \qquad (11)$$

and then Φ is integer modulo N.

Thus notice that the dual similarity between electric and magnetic fields is still there in the non-Abelian theory. They both are conserved modulo N, and must be integers. We further note that the operators A(C) and B(C') do not commute (just as in the Abelian case). One finds:

$$A(C)\, B(C') = B(C')\, A(C)\, e^{2\pi i n/N} , \qquad (12)$$

where n is now the number of times that the curve C' winds through C.

THE HIGGS MODEL: NIELSEN-OLESEN VORTICES

Properties of the operators A and B can be studied in a soluble model. With soluble I mean that perturbation expansion can be applied and converges sufficiently rapidly so that the results can be trusted. Pure gauge theories are not (yet?) soluble in this sense because of infrared divergences. But we can add scalar field multiplets H_i that transform under gauge transformations the way the Hermitean matrices A_μ do:

$$H_i' = \Omega\, H_i\, \Omega^{-1} . \qquad (13)$$

And choose the self-interactions of H such that for instance

$$\langle H_i \rangle_{vacuum} = F_i\, I \neq 0 . \qquad (14)$$

The local gauge invariance is then spontaneously broken. All particles become massive so that infrared divergencies disappear.

The effect of the operator A(C) in this model is easily understood when we expand the exponent. The expansion converges by assumption and we find that the operator creates gauge quanta along the curve C. The vacuum expectation value of A(C) is obtained through a Feynman diagram technique where photon propagators are attached both ends onto the curve C. One finds, when C becomes large, that

$$\langle A(C) \rangle_{vacuum} \to \alpha_1 \exp(-\alpha_2\, L(C)) \qquad (15)$$

where L(C) is the total length of C.

The effect of B(C) is entirely different. The model not only contains gauge quanta and Higgs particles but also a locally stable vortex structure, called Nielsen-Olesen vortex or string[7]. The stability of this string is guaranteed by the boundary condition: far away from the vortex we require

$$A_\mu \rightarrow \frac{1}{gi}\, \Omega\, \partial_\mu\, \Omega^{-1} \tag{16}$$

where Ω is multivalued when we follow it around the vortex. We observe that B(C) creates just this new boundary condition, so a Nielsen-Olesen vortex is created that coincides with the curve C. Now the physically stable Nielsen-Olesen vortex has a finite width. What B(C) creates is an "excited state" of this vortex that carries more energy per unit of length but whose transverse dimensions start off very small. It will soon decay into the stable vortex plus some gauge quanta. The stable vortex, when created along a closed contour, will finally also decay, but only by shrinking the complete contour defined by its transverse center of gravity. One can show that therefore for large closed curves C,

$$\langle B(C)\rangle_{\text{vacuum}} \rightarrow \beta_1 \exp - \beta_2\, \Sigma(C) \tag{17}$$

where Σ(C) is the total area spanned by C.
Summarizing: because the gauge quanta are massive the potential between two electric charges decays exponentially with distance, but because the magnetic vortex carries energy per unit of length, the potential between two elementary magnetic charges rises linearly with distance.

A PHASE TRANSITION. THE VACUUM AS SUPER-INSULATOR FOR QUARKS

The way the operators A and B behave in the well-known Higgs model is opposite to what we expect in a quark theory. There it is the *electric* potential which we expect to rise linearly with distance. We now argue that, because the operators A and B are so similar in nature, a phase transition is possible that interchanges the long-distance characters of A and B, if we vary the parameters in the theory. In fact we expect this other phase to be realized also in a model without Higgs fields. What can be proven[5] is the following. Because of the commutation rules (12) the possible ways in which the observables A and B behave for large curves C, C' are restricted. One can have either

Higgs mode:

$$\begin{cases} \langle A(C)\rangle \rightarrow \alpha_1 \exp(-\alpha_2\, L(C)) \\ \langle B(C)\rangle \rightarrow \beta_1 \exp(-\beta_2\, \Sigma(C)) \end{cases}$$

or
Confinement mode:

$$\begin{cases} \langle A(C)\rangle \to \alpha_1 \exp(-\alpha_2\ \Sigma(C)) \\ \langle B(C)\rangle \to \beta_1 \exp(-\beta_2\ L(C)) \end{cases}$$

or, but unlikely, both Higgs and confinement:

$$\begin{cases} \langle A(C)\rangle \to \alpha_1 \exp(-\alpha_2\ \Sigma(C)) \\ \langle B(C)\rangle \to \beta_1 \exp(-\beta_2\ \Sigma(C)) \end{cases}$$

or an infrared divergent mode with massless particles, presumably representing the critical point where the phase-transition occurs. One cannot have

$$\begin{cases} \langle A(C)\rangle \to \alpha_1 \exp(-\alpha_2\ L(C)) \\ \langle B(C)\rangle \to \beta_1 \exp(-\beta_2\ L(C)) \end{cases}$$

corresponding to a state with neither Nielsen-Olesen vortex nor quark confining strings.
The Higgs model can be regarded as a non-Abelian generalization of a super conductor with the Nielsen-Olesen vortex describing the Meissner effect.
The confinement mode is the dually opposite of a super conductor. The vacuum is a super insulator for quarks.

Since the flux conservation law only holds modulo N, the model will allow N quarks to sit together with finite total energy. Baryons are observed to be composed of 3 quarks. Apparently for them N = 3.

OPEN QUESTIONS AND PROBLEMS

The theory sketched above can be worked out much more quantitatively but still there are many problems to be solved. First, we did not prove that quark confinement occurs, only that it is one of the few possible modes that are separated by phase transition points. It is extremely hard, in fact not yet possible, to do reliable calculations in a pure gauge theory with fermions. We do not know how to compute baryon and meson mass spectra reliably, let alone their cross sections. We believe that the Lagrangean that is written down for the model does indeed describe everything (apart from a by now well understood vacuum degeneracy[8]) but we have no mathematical proof of this, so it could be wrong.

Secondly, the model does have some (though few) degrees of freedom. There is a free parameter g, fixing the mass scale, and then there is a free parameter corresponding to the "mass" of each quark multiplet added. The total number of quark multiplets is free as long as it does not exceed 16. At present the number is 5. If it does exceed 16 then definitely new physics must come in at very small distances.
Are these parameters really free? is there no closer link between strong and weak interactions? We do not know. There are many specula-

lations on unification of strong and weak interactions, but these give so many possibilities that the answer might have to depend on either super energetic accelerators or fundamentally new theoretical ideas.

REFERENCES

1. C.N. Yang and R.L. Mills, Phys. Rev. 96, 191 (1954).
2. R. Shaw, Cambridge Ph.D. Thesis, unpublished.
3. J. Kogut and L. Susskind, Phys. Rev. D9, 3501 (1974).
   K.G. Wilson, Phys. Rev. D10, 2445 (1974).
4. H.D. Politzer, Phys. Rev. Lett. 30, 1346 (1973).
   D.J. Gross and F. Wilczek, Phys. Rev. Lett. 30, 1343 (1973).
5. G. 't Hooft, Utrecht preprint, to be published in Nucl. Phys. B.
6. F. Englert and R. Brout, Phys. Rev. Lett. 13, 321 (1964).
   P.W. Higgs, Phys. Lett. 12, 132 (1964), Phys. Rev. Lett. 13, 508 (1964), Phys. Rev. 145, 1156 (1966).
7. H.B. Nielsen and P. Olesen, Nucl. Phys. B61, 45 (1973).
   B. Zumino, in Renormalization and Invariance in Quantum field theory, ed. by E.R. Caianiello (1974 Plenum Press, New York).
   Y. Nambu, Phys. Rev. D10, 4262 (1974).
8. S. Coleman, Lectures delivered at the 1977 Int. School of Subnuclear Physics, Ettore Majorana.