GAUGE THEORY FOR STRONG INTERACTIONS

G. 't Hooft

Institute for Theoretical Physics, University of Utrecht, Sorbonnelaan 4, Utrecht, The Netherlands

1. INTRODUCTION

There are good indications that the strong interactions can be described by a non-Abelian gauge theory based on the "color" gauge group SU(3). There are only gauge vector bosons and quarks in this theory, but no scalar fields, hence we have no symmetry breaking through a Higgs mechanism.

This theory has the very important property that it is asymptotically free. This means the following: Our elementary quanta are in principle massless, so we can choose different reference mass scales to describe scattering amplitudes. Going from one reference mass to a larger reference mass is like looking at the system through a microscope. What we then see is the same set of elementary quanta, but with a smaller interaction constant. In the limit of infinite magnification, we finally see free particles. Not only does this correspond to the hadron structure that is observed experimentally, but also from a theoretical, mathematical, point of view this is a very desirable feature which will perhaps enable us to put such a strong interaction theory on a mathematically sound footing.

Thus the ultraviolet behaviour of the theory looks good, but, conversely, the infrared problems are formidable. Suppose we consider the system in a box with sides of length L. For small L the physical degrees

of freedom are still described by the elementary quanta, but for larger L the effective interaction strength increases rapidly, and if L is much larger than a certain value L_o then perturbation expansions can no longer be trusted. Besides the elementary quanta there will be many bound states. It is therefore very difficult to understand the limit $L \to \infty$.

There are essentially two possibilities for the $L \to \infty$ theory: the first is that the higher order corrections to the Callan-Symanzik β -function may show an infrared fixed point g_o. Then the limiting theory has a continuous particle spectrum with long range strongly interacting fields, and is scale invariant. This has clearly nothing in common with our real world. We therefore believe that a second possibility is realized: the $L \to \infty$ theory has a new mass spectrum which has no massless particles. The lightest particles have a mass of the order of $1/L_o$.

The elementary quanta (quarks, gauge vectorbosons) are now the constituents of the physical particles (the hadrons). Let us now argue why we think that these hadrons can only consist of quark-antiquark pairs or triplets of three quarks[1]. Let us consider the subset of Hilbert space that contains all states with one quark and one antiquark pinned at fixed positions, and let us estimate the energy of the lowest-energy state in that subspace, as a function of the seperation of these quarks. We expect a color-electric field pointing from one quark to the other, but it cannot possibly have the same configuration as ordinary Maxwell fields have (fig. 1)

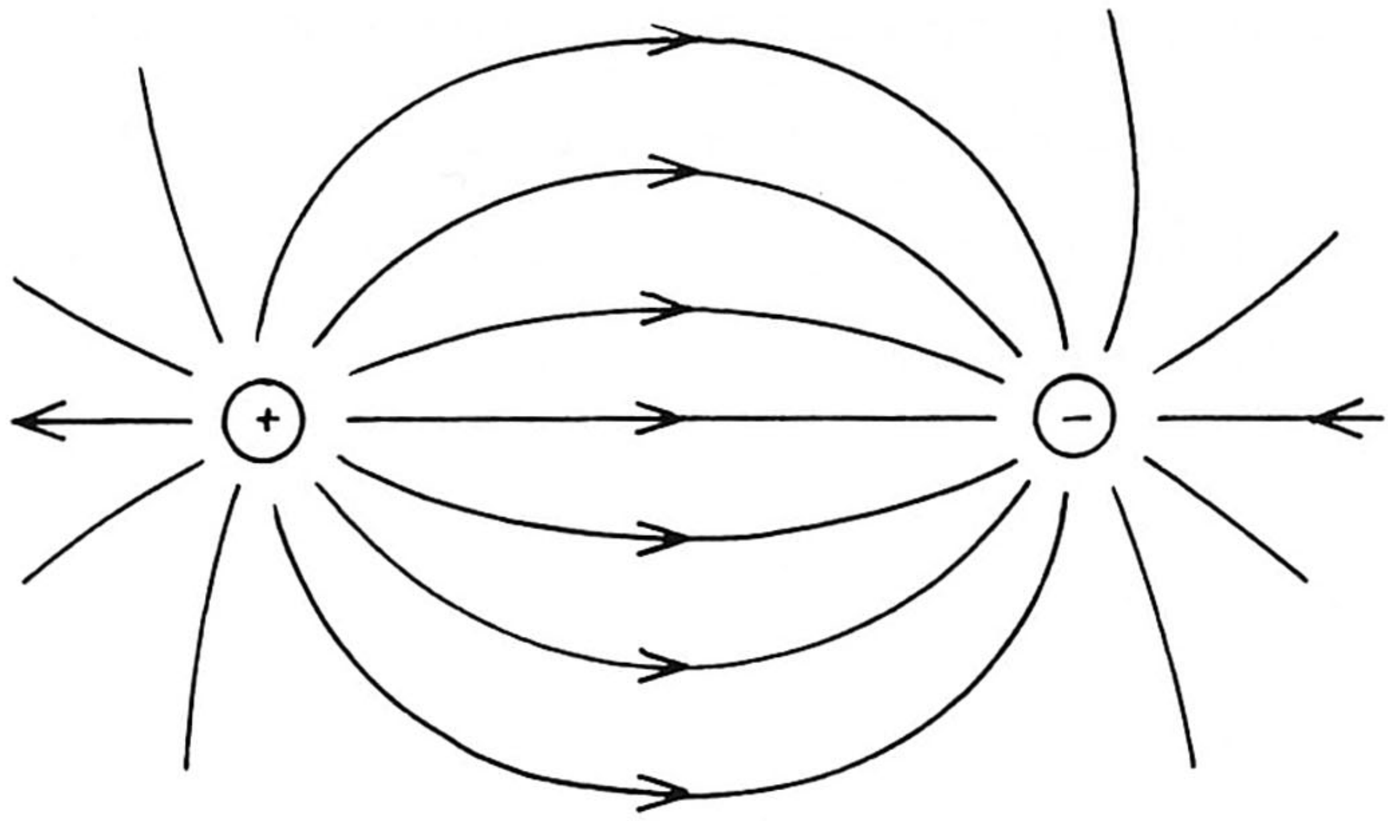


fig. 1. Maxwell fields between two charges.

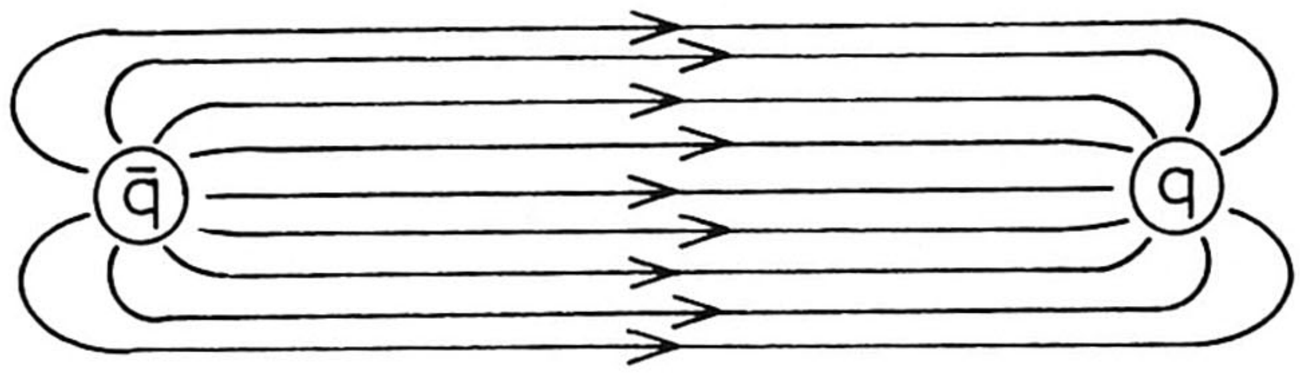


fig. 2. Color electric field between a quark-antiquark pair.

Such fields would be described by linear equations with no mass terms, so there would be massless quanta, contrary to our assumption. We rather expect that, as a function of the distance r to the symmetry axis, we would get a decrease like e^{-mr} where m is the smallest physical hadronic mass. If color-flux is conserved, that would result in a configuration like in fig. 2.

The force between the quark-antiquark pair would be equal to the energy of this "tube" per unit of length, and thus independent of the distance, as soon as the distance is larger than $1/m$. It is this way that we imagine how a dynamical effect can confine any color non singlet object.

In the following sections we show how several approximation schemes help us to obtain a consistent picture of this color confinement mechanism. It has been proposed by Fritzsch, Gell-Mann and Minkowsky to call this theory "Quantum Chromodynamics" (QCD).

2. FEYNMAN RULES FOR U(N) GAUGE THEORIES

We now proceed by making a generalization: we replace color SU(3) by SU(N). All arguments we used for confinement, so far, should hold for all $N \geqslant 2$. Our theory must be a decent theory for all positive integer N. Only the baryons change in character: they are now built from N quarks instead of three. This is because in SU(N) one can make invariant, completely antisymmetric ε - tensors with N indices. Color singlets are therefore

$$\varepsilon_{i_1 \ldots i_N} \; q_{i_1} \cdots q_{i_N} \; . \qquad (2.1)$$

But we will concentrate on mesons.

Next, we make a simplification: we replace the color group SU(N) by U(N). That merely corresponds to adding an Abelian field coupled with the same strength to "quark number", as the non-Abelian field to color. Such a photon will certainly not be able to confine anything, and we can always change its coupling strength in the end, for instance by cancellation with a new photon. In any case this photon will not affect the leading behaviour at $N \to \infty$, as we can show. We will disregard it from now on.

Our gauge vector fields are antihermitean matrices:

$$W_i{}^j{}_\mu(x) = -W^{*\,i}_j{}_\mu(x) \qquad (2.2)$$

$$(\mu = 0,1,2,3 \;\; ; \;\; i,j = 1,\ldots,N).$$

We do not make the restriction that they are traceless:

$$W_i{}^i{}_\mu \neq 0. \qquad (2.3)$$

The Lagrangian is

$$\mathcal{L} = +\tfrac{1}{4} G_{\mu\nu\,i}{}^j G_{\mu\nu\,j}{}^i - \bar{q}^{ai}(\gamma_\mu D_\mu + m^{(a)}) q^a{}_i \;, \qquad (2.4)$$

where*)

$$q^1{}_i = P_i \;\; ; \;\; q^2{}_i = n_i \;\; ; \;\; q^3{}_i = \lambda_i \;\; ;$$

$$G_{\mu\nu\,i}{}^j = \partial_\mu W_i{}^j{}_\nu - \partial_\nu W_i{}^j{}_\mu + g[W_\mu, W_\nu]_i{}^j \;\; ; \qquad (2.5)$$

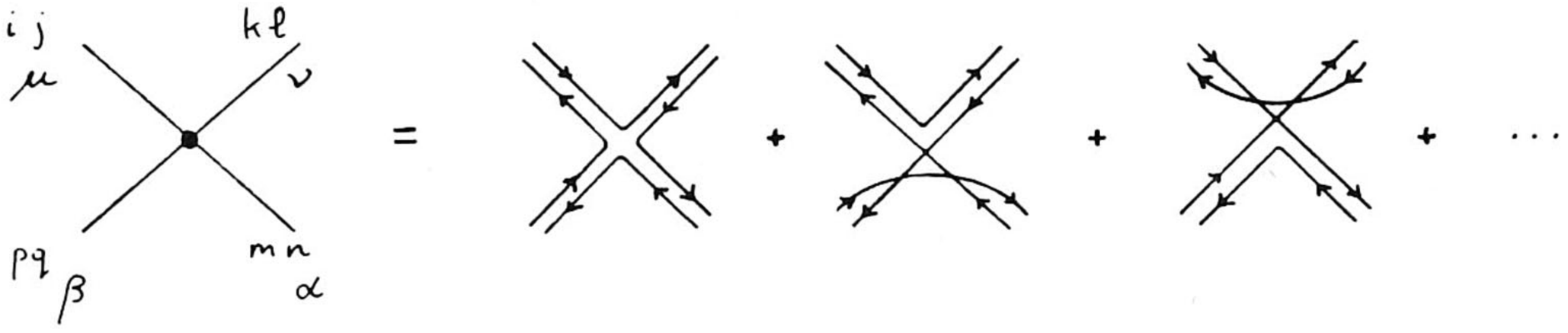
$$D_\mu q^a{}_i = \partial_\mu q^a{}_i + g W_i{}^j{}_\mu \, q^a{}_j \;.$$

The unusual + sign in (2.4) is a consequence of the antihermitean nature of the fields (2.2).

The Feynman rules are now derived in a completely standard way[2]. The propagators are minus the inverse of the bilinear parts in the Lagrangian, and the vertices are derived from the interaction parts. In what follows we want to keep track of the dependence of the amplitudes on the color group U(N). Therefore we invent a special notation for each term in the vertices that depends differently on the color-indices. The color-indices only combine in Kronecker delta functions, in our notation. Each gauge field line carries two color indices, the fermion lines just one. We now say that for instance the four point vertices contain six terms (fig. 3).

---

*) Note that some numerical coefficients may differ from more conventional notations. They have been chosen to get simple Feynman rules for general



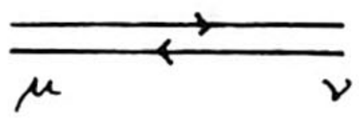
$$V^{ikmp}_{j\ell nq}\big|_{\mu\nu\alpha\beta} = \delta^k_j \delta^m_\ell \delta^p_n \delta^i_q V_{\mu\nu\alpha\beta} +$$
$$+ \delta^k_j \delta^p_\ell \delta^i_n \delta^m_q V_{\mu\nu\beta\alpha} + \delta^m_j \delta^i_\ell \delta^p_n \delta^k_q V_{\mu\alpha\beta\nu} + \cdots$$

fig. 3

In fig. 3 we denoted the color indices explicitly. Two indices that occur in one Kronecker delta are connected by one line. The lines have arrows to distinguish upper indices from lower indices. We observe that the Lorentz index structure of all these terms is described by just one symbol $V_{\mu\nu\alpha\beta}$, because of Bose symmetry. The Feynman rules are now determined as soon as we give the object V belonging to the first structure in fig. 3. The same is true for the other vertices. In fig. 4 we give all Feynman rules. There may be a Faddeev-Popov-Feynman-DeWitt ghost, depending on the gauge chosen, but it is irrelevant for our present considerations.

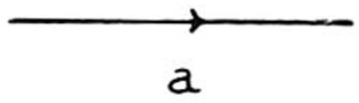
3. THE LARGE N LIMIT

How do the amplitudes depend on the parameter N? Clearly, in our set-up, the Feynman rules do not explicitly depend on N, except for the possible values the color indices may take. But now let us compare two possible diagrams (fig. 5a and b).



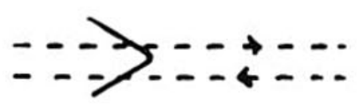
color gauge
vector boson

$$-\frac{\delta_{\mu\nu}}{k^2 - i\varepsilon}$$



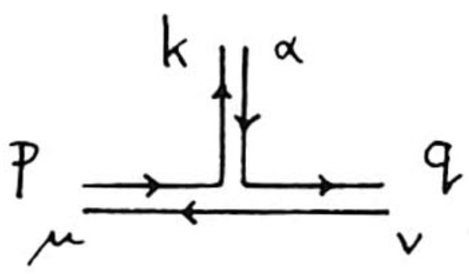
a = 1 p quark
a = 2 n quark
a = 3 λ quark

$$\frac{1}{m_i^{(a)} + i\gamma k - i\varepsilon}$$



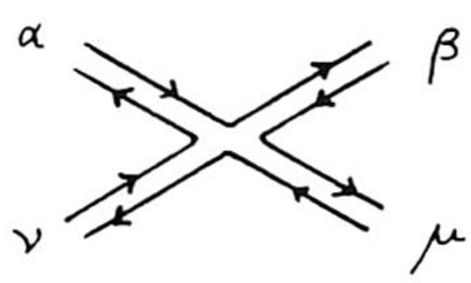
ghost

$$\frac{1}{k^2 - i\varepsilon}$$



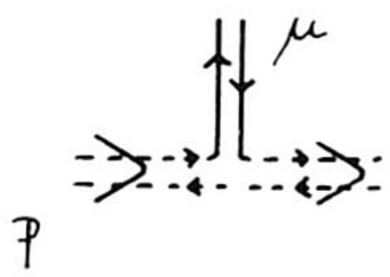
vertices

$$ig\{\delta_{\alpha\nu}(k-q)_\mu + \delta_{\alpha\mu}(p-k)_\nu + \delta_{\mu\nu}(q-p)_\alpha\}$$



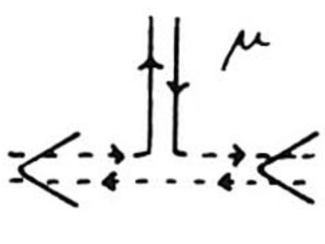
"

$$g^2\{2\delta_{\alpha\mu}\delta_{\beta\nu} - \delta_{\alpha\beta}\delta_{\mu\nu} - \delta_{\beta\mu}\delta_{\alpha\nu}\}$$



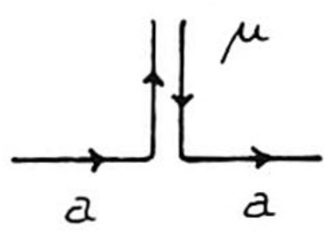
"

$$igp_\mu$$



"

$$-igp_\mu$$



"

$$-g\gamma_\mu$$

each closed loop of ghosts: -1

fig. 4

The Feynman rules for general N, in the Feynman gauge.

a b

fig. 5

We observe that an index-line can make a closed loop. Such a loop gets a factor

$$\delta_i^{\ i} = N. \tag{3.1}$$

The first diagram contains two such loops. So its color factor is N^2. But the second has only one index loop. So it gets a factor N only. The Lorentz structure and the momentum-dependence of both diagrams are independent of N. So clearly, if we take N large enough, we may neglect diagram b with respect to diagram a.

Diagram a) is completely planar whereas diagram b) cannot be drawn without crossings (if we insist that the external lines point outwards). Is it accidental or is it a general rule that those diagrams that can be drawn on a plane without crossings have more factors of N than those which do not fit on a plane? The answer is that this indeed is a general rule and we shall prove that now[3,4].

It is more convenient to limit ourselves to vacuum expectation values of gauge invariant sources so we imagine source terms like

$$\jmath_1 \bar{P}P + \jmath_2 \bar{P}n + \jmath_3 \bar{n}\gamma^5\lambda + \text{etc.}$$

in the Lagrangian. They describe insertions like fig. 6 in the diagrams.

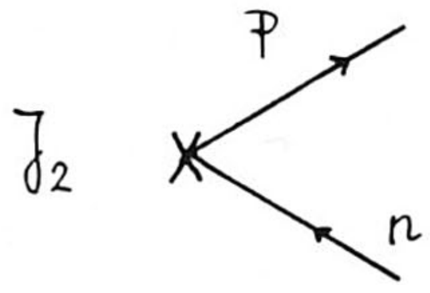


fig. 6. Source insertions.

It is important to note that we must always sum over the color-indices, so again, if the line in fig. 6 forms an index-loop, then that gives a factor N. All diagrams now necessarily also contain Fermion lines which are forced to make complete loop(s).

Consider now an arbitrarily complicated diagram that has no external lines but only source insertions. We will call the latter "two point vertices". Now let the number of two- , three- and four point vertices be resp. V_2 , V_3 and V_4 ,

and let $V = V_2 + V_3 + V_4$

Let there be P propagators,

I index loops,

and F Fermion loops.

Our first theorem is

$$2P = 2V_2 + 3V_3 + 4V_4 \quad . \tag{3.2}$$

This we prove, by imagining a dot at each end of each propagator. Then we can count the total number of dots either by counting the propagators or by counting the vertices. This must yield the same number.

Next we define the topology of a diagram by making a two-dimensional surface out of it, in the following way: each index loop will be the edge of a singly connected surface, which we will call a "window". Each gauge field propagator will glue together two windows, and so we get a big surface, with perhaps many wormholes. Each fermion loop

will give an ordinary hole in the surface, because fermion lines just make an edge to each window. We also close these holes by means of windows. So the total number of windows is $I + F$.

We now have a closed and oriented surface. Its topology is simply defined by counting the number H of wormholes or handles. We can then apply a theorem due to Euler:

$$I + F + V - P = 2 - 2H \quad . \tag{3.3}$$

This is easily proved, first if there are no wormholes, and then by means of induction with respect to the number of wormholes.

The two free parameters of our field theory are g and N. The factor that associates our big diagram is

$$\begin{aligned} & g^{V_3+ 2V_4}\, N^I \\ &= g^{2P-2V}\, N^I \\ &= (g^2 N)^{P-V}\, N^{2-2H-F} \\ &= \tilde{g}^{V_3+2V_4}\, N^2 \left(\tfrac{1}{N}\right)^{2H} \left(\tfrac{1}{N}\right)^{F} \quad , \end{aligned} \tag{3.4}$$

where $\tilde{g} = g\sqrt{N}$.

Now $H \geqslant 0$ and $F > 0$, so if we keep $\tilde{g}$ fixed, and let $N \to \infty$, then the dominating diagrams are those with $H = 0$ and $F = 1$. There is an overall normalization factor N. This proves that the dominating diagrams fit on a single plane with no wormholes and just one complete Fermion loop at its edge. An example is pictured in fig. 7.

The subsequent figure shows the next order diagrams in 1/N. Fig. 8a has one factor 1/N, and fig. 8b one factor $1/N^2$. It is remarkable that we obtain exactly the same topological structures as in the dual loop

expansion in dual theories[5]. Our coefficient 1/N evidently corresponds to the expansion parameter λ in dual theories. For $\lambda \to 0$ the resonances in dual theories move towards the real axis; their interactions vanish. In our theory, if $N \to \infty$, then the planar diagrams of fig. 7 cannot be cut in such a way that we get more than one color singlet. So if the theory behaves as we think, then also in our theory mesons become stable as $N \to \infty$.

The importance of the 1/N expansion is perhaps not that the physical value of the expansion parameter is 1/3 which is rather small, but that the $N \to \infty$ limit theory already seems to have a complete spectrum of mesons. The 1/N corrections will probably be unimportant and will not be plagued by infrared troubles. Of course it has the disadvantage that we do not see the baryons this way. The fact that mesons do not interact in the limit $N \to \infty$ can also be proved in lattice gauge theories.

4. TWO DIMENSIONS

It will be evident from figs. 7 and 8 that we are unable to sum all planar diagrams in this field theory, so we will be forced to make more approximations or simplifications. One such simplification is to consider the theory in only one space- and one time dimension. This sounds drastic but the results are very interesting and pedagogically important.

In two dimensions we can choose the gauge

$$a_\mu W_\mu = 0 \quad , \tag{4.1}$$

with a_μ any non-vanishing vector. Then the Lorentz vector W_μ has only one non-vanishing component, which of course commutes with itself. The interaction term in $G_{\mu\nu}$, eq. (2.5), which is a commutator, therefore vanishes and so all gauge field self-interactions are zero. The set of planar diagrams, for a given channel, then have the form of fig. 9.

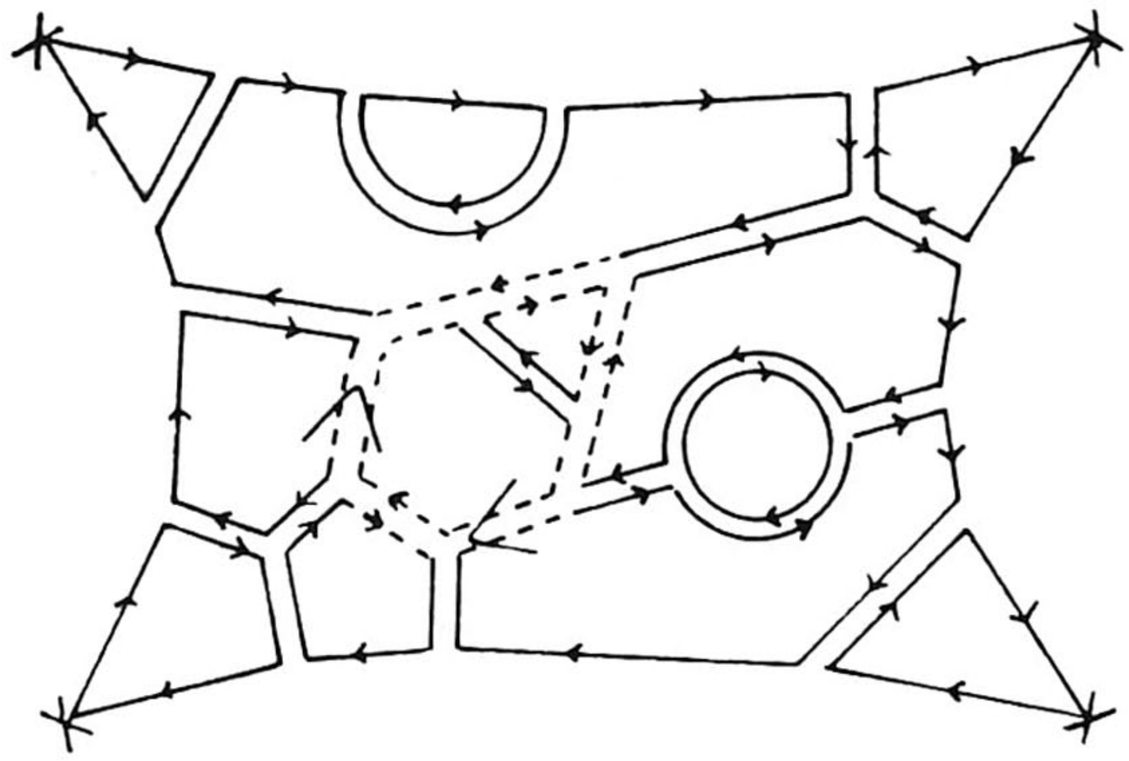
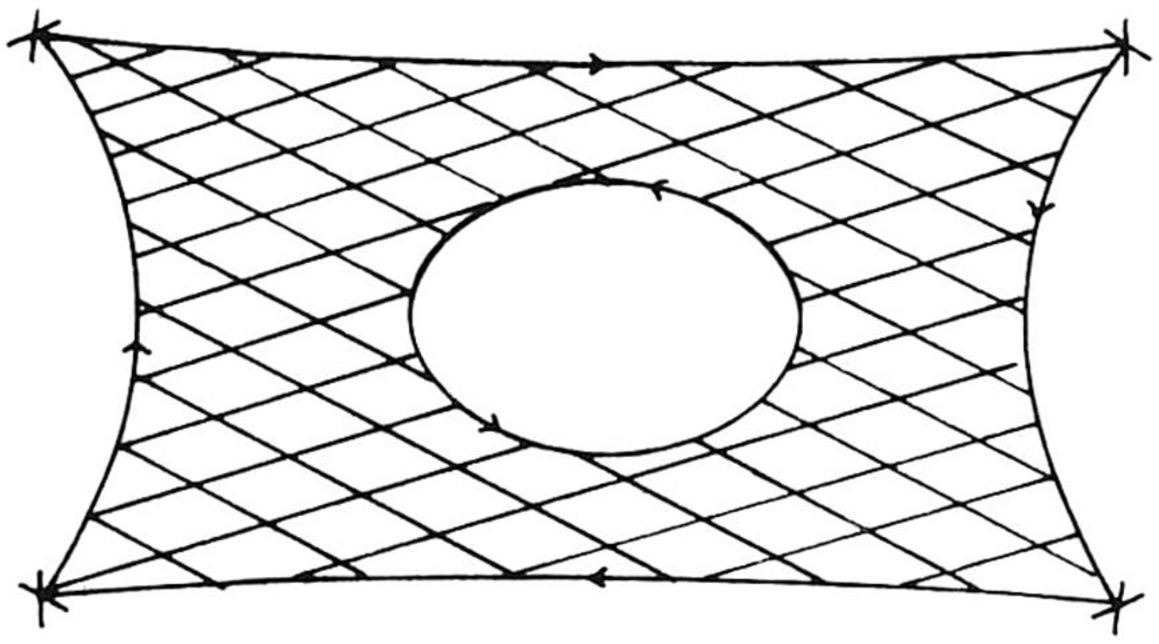
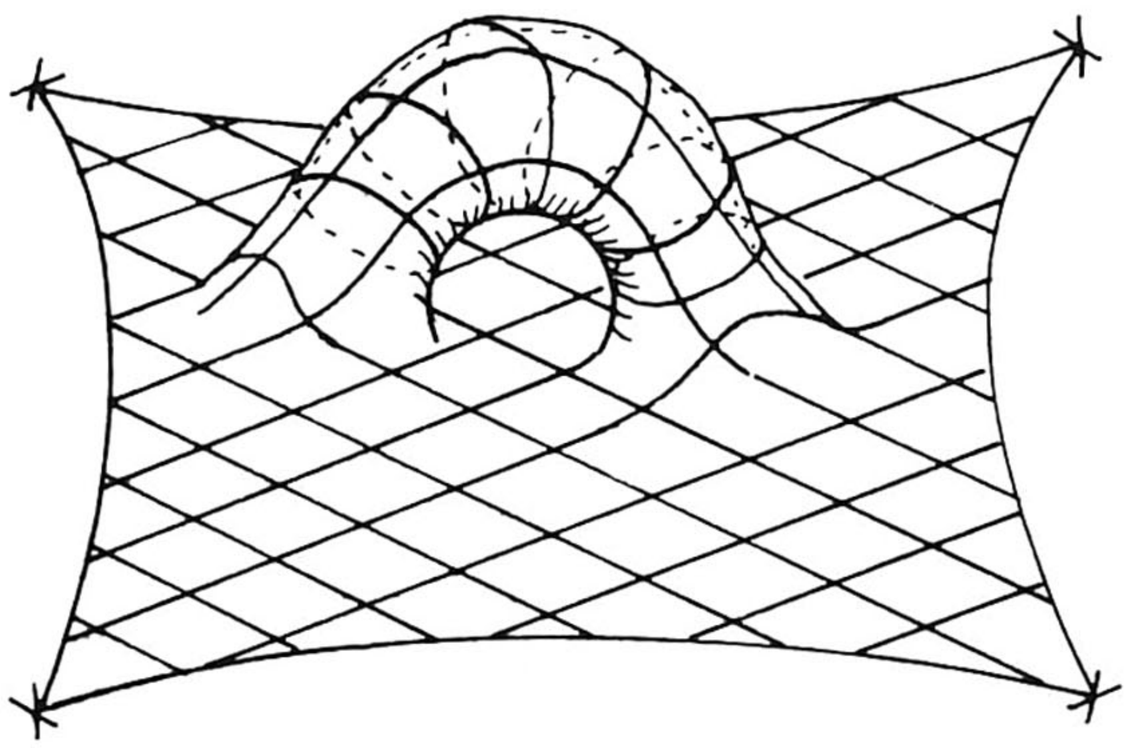


fig. 7.



a.



b.

fig. 8.

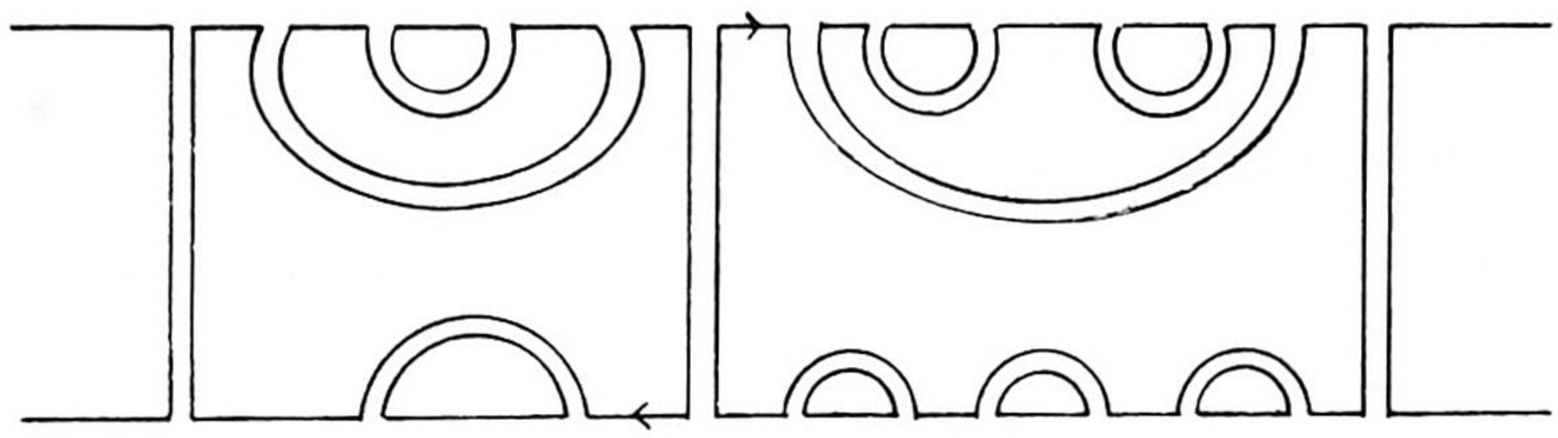


fig. 9. The planar diagrams in two dimensions.

This set is exactly summable[6], but still rather complicated unless we choose the light cone gauge:

$$a^2 = 0, \quad \text{or} \quad W_- = 0 \ . \tag{4.2}$$

For the details of the calculations we refer the reader to ref.(6). Here we will merely quote the results. We will use the notation:

$$x^{\pm} = \frac{1}{\sqrt{2}}(x^1 \pm x^0) \qquad \text{for upper indices,}$$

and

$$p_{\pm} = \frac{1}{\sqrt{2}}(p_1 \pm p_0) \ ,$$

$$W_{\pm} = \frac{1}{\sqrt{2}}(W_1 \pm W_0) \ , \text{ etc.} \qquad \text{for lower indices,}$$

where

$$p_1 = p^1, \qquad p_0 = -p^0 .$$

Our summation convention will be

$$x_\mu p^\mu = x^\mu p^\mu = x_\mu p_\mu =$$

$$x_+ p^+ + x_- p^- = x^+ p^- + x^- p^+ = x_+ p_- + x_- p_+ \ .$$

The Lagrangian is now

$$\mathcal{L} = -\frac{1}{2} \mathrm{Tr} \left(\partial_- W_+\right)^2 - \bar{q}^a(\gamma\partial + m^{(a)} + g\gamma_- W_+) q^a \ , \tag{4.3}$$

and there is no ghost in this gauge.

The system is infrared divergent so we must think of a reasonable cutoff that makes physically sence. The best cutoff is to put the system in a box with finite size L and to consider the limit $L \to \infty$ in the end. Our boundary conditions are the periodic ones. Let us first do this in the radiation gauge:

$$\begin{pmatrix} a_1 \\ a_o \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad . \tag{4.4}$$

This gauge is singular at zero momentum, so we must be more accurate:

$$W_1 (k \neq 0) = 0 \quad , \tag{4.5a}$$

$$W_o (k = 0) = 0 \quad . \tag{4.5b}$$

In (4.5a) we must put $k \neq 0$ because the zero momentum part of W_1 is gauge invariant to lowest order:

$$\int_L \frac{\partial}{\partial x_1} \Lambda(x) = 0 \quad ,$$

since Λ also must be periodic. We may choose (4.5b) because (4.5a) still leaves one invariance: $\Lambda(x,t) = \Lambda(t)$.

From (4.5b) we get the revised photon propagator

$$\frac{1}{k_1^2} (1 - \delta_{k_1 0}) \quad , \tag{4.6}$$

whereas the zero momentum part of W_1 gives rise to a new interaction which however becomes of zero measure as $L \to \infty$.

The net effect of this procedure is that we may cut off the singularity at $k_1 \to 0$. We could just as well keep the momenta continuous, and replace $1/k_1^2$ by

$$\frac{1}{k_1^2} \theta(k_1^2 - \lambda^2) \quad , \tag{4.7}$$

where λ is now the effective cutoff. The Coulomb potential is then the

Fourier transform of (4.7), pictured in fig. 10. So we see that this cutoff prescription can be interpreted in two ways: it is obtained when the system is put in a box, or when alternatively the Coulomb potential is cut off as in fig. 10.

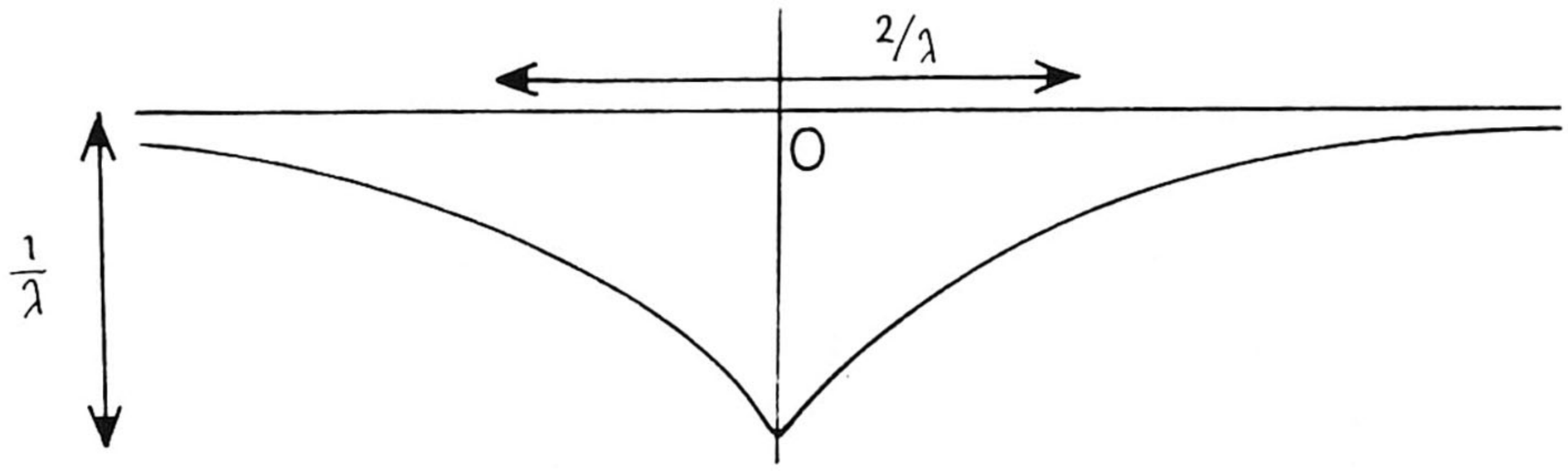


fig. 10. The Coulomb potential with cutoff.

For the light cone gauge, which is just the limit of an infinite Lorentz boost of the radiation gauge, we may use the same cutoff procedure:

$$\frac{1}{k_-^2} \rightarrow \frac{1}{k_-^2}\,\theta(k_-^2 - \lambda^2)\ . \tag{4.8}$$

With this cutoff prescription the dressed quark propagator is

$$\frac{-ik_-\gamma_+}{m^2 - \frac{g^2N}{\pi} + 2k_+k_- + g^2N\frac{|k_-|}{\pi\lambda} - i\varepsilon}\ . \tag{4.9}$$

For the details of this calculation see ref. 6*).

We see that there is a Lorentz non-invariant infinite self energy:

$$g^2N\frac{|k_-|}{\pi\lambda}\ . \tag{4.10}$$

This is Lorentz non-invariant because the box introduces a preference

---

*) In ref. 6, from eq. 10 onwards, g^2 should read g^2N.

frame. We will see that colorless states become cutoff independent and Lorentz invariant.

The diagrams of fig. 9 are then summed by solving the Bethe-Salpeter equation with the dressed propagators. If

$$\Psi(p_+ \cdot p_- \cdot r)\ \gamma_+(1)\ \tilde{\gamma}_+(2) \tag{4.11}$$

stands for the wave function for a quark (1) with momentum p and an antiquark (2) with momentum r - p then we find that the relevant function is

$$\varphi(p_-, r) = \int \Psi(p_+, p_-, r)\ dp_+ \tag{4.12}$$

which means that we only have to consider the quarks at equal time ($x^+(1) = x^+(2)$).

We take $r_- > 0$ and find that

$\varphi = 0$ if $p_- < 0$ or $p_- > r$, and otherwise:

$$\varphi(p_-,r) = -\frac{g^2N}{(2\pi)^2 i}\left\{\int dp_+\left[p_+ - r_+ + \frac{M_2^2}{2(p_- - r_-)} + \left(\frac{g^2N}{2\pi\lambda} - i\varepsilon\right)\mathrm{sgn}(p_- - r_-)\right]^{-1}\right.$$
$$\left.\left[p_+ + \frac{M_1^2}{2p_-} + \left(\frac{g^2N}{2\pi\lambda} - i\varepsilon\right)\mathrm{sgn}(p_-)\right]^{-1}\right\}\int dk_-\ \frac{\varphi(p_- + k_-, r)}{k_-^2}\,\theta(k_-^2 - \lambda^2) = \tag{4.13}$$
$$= \frac{g^2N}{2\pi}\left[\frac{M_1^2}{2p_-} + \frac{M_2^2}{2(r_- - p_-)} + \frac{g^2N}{\pi\lambda} + r_+\right]^{-1}\int dk_-\ \frac{\varphi(p_- + k_-, r)}{k_-^2}\,\theta(k_-^2 - \lambda^2).$$

Here $M_i^2 = m_i^2 - g^2N/\pi$.

The integral with the λ dependent cutoff can be replaced by a principle-value integral:

$$\int dk_-\ \frac{\varphi(p_- + k_-, r)}{k_-^2}\,\theta(k_-^2 - \lambda^2) =$$
$$= \frac{2}{\lambda}\varphi(p_-) + \mathcal{P}\int \frac{\varphi(p_- + k_-, r)}{k_-^2}\,dk_- + O(\lambda). \tag{4.14}$$

The integration is now over a contour in the complex plane that avoids the pole, and we average over the two possible ways we can go around the pole. We now find that the cutoff dependent terms cancel, and

$$-r_+ \varphi(p_-, r) = \left(\frac{M_1^2}{2p_-} + \frac{M_2^2}{2(r_- - p_-)} \right) \varphi(p_-, r) - \frac{g^2 N}{2\pi} P \int_{-p_-}^{r_- - p_-} \frac{\varphi(p_- + k_-, r)}{k_-^2} dk_- . \quad (4.15)$$

Introducing

$$\begin{aligned} p_-/r_- &= x \ , \\ -2r_+ r_- &= \mu^2 \frac{g^2 N}{\pi} \ , \\ m_{1,2}^2 &= (\alpha_{1,2} + 1) \frac{g^2 N}{\pi} \ , \end{aligned} \quad (4.16)$$

the equation becomes

$$\mu^2 \varphi(x) = \left(\frac{\alpha_1}{x} + \frac{\alpha_2}{1-x} \right) \varphi(x) - P \int_0^1 \frac{\varphi(y)}{(y-x)^2} dy . \quad (4.17)$$

At the points $x = 0$ and $x = 1$ the function φ must be zero (for $m_i^2 > 0$) or finite (for $m_i^2 = 0$). This follows from the requirement that the operator at the right hand side of eq. (4.17) be hermitean[6].

5. THE MASSES OF THE BOUND STATES

Our integral equation (4.17), with the correct boundary conditions, gives a perfectly well behaved spectrum of bound states. We may rewrite (4.17) as

$$\mu^2 \varphi(x) = \left(\frac{\alpha_1 + 1}{x} + \frac{\alpha_2 + 1}{1-x} \right) \varphi(x) - P \int_0^1 \frac{\varphi(y) - \varphi(x)}{(y-x)^2} dy \ , \quad (5.1)$$

from which it follows that

$$\begin{aligned} \mu^2 \int_0^1 \varphi^* \varphi \, dx &= \int_0^1 \left(\frac{\alpha_1 + 1}{x} + \frac{\alpha_2 + 1}{1-x} \right) \varphi^* \varphi \, dx + \\ &+ \frac{1}{2} \int_0^1 dx \int_0^1 dy \, \frac{[\varphi^*(x) - \varphi^*(y)][\varphi(x) - \varphi(y)]}{(x-y)^2} \ , \end{aligned} \quad (5.2)$$

so $\mu^2 \geqslant 0$, and $\mu^2 = 0$ is only possible if $m_1^2 = m_2^2 = 0$. This is the only case in which we could solve this equation: if $m_1^2 = m_2^2 = 0$ the ground state is

$$\varphi_o(x) = 1 \tag{5.3}$$

A first approximation for the very highly excited states is given by a rapidly oscillating function,

$$\varphi_k(x) \simeq \sin k\pi x , \qquad k = 1, 2, \cdots , \tag{5.4}$$

for which we have

$$\begin{aligned} P\int_0^1 \frac{\sin k\pi y}{(y-x)^2}\, dy \;&\simeq P\int_{-\infty}^{\infty} \frac{\sin k\pi y}{(y-x)^2}\, dy \;= \\ &= -\pi^2 |k| \sin k\pi x . \end{aligned} \tag{5.5}$$

Taking $k \gg \frac{\alpha_1}{x} , \frac{\alpha_2}{1-x}$,

we get

$$\begin{aligned} (\mu^2 - \pi^2|k|)\sin k\pi x \;&\simeq 0 , \\ \mu^2 \underset{k\to\infty}{\longrightarrow} \;&\pi^2 k . \end{aligned} \tag{5.6}$$

This is a straight "Regge" trajectory. A more precise WKB approximation yields

$$\mu^2 \longrightarrow \pi^2 k + (\alpha_1 + \alpha_2)\log k + C^{st}(\alpha_1, \alpha_2) . \tag{5.7}$$

If we put the units back in, we get that the slope of the trajectories is

$$\alpha' = 1/\pi g^2 N . \tag{5.8}$$

Of course we must keep in mind that we do not really have Regge trajectories but rather series of daughters. The higher states have no angular momentum, but they are radial excitations.

Numerical calculations show that in the case $\alpha_1 = \alpha_2 = 0$ a straight line is approached rapidly, with an intercept $\alpha(0) = -3/4$.

Let us now consider the case

$$m_1^2 , m_2^2 \ll \frac{g^2 N}{\pi} , \tag{5.9}$$

and see what the first corrections are to the zero-mass bound state (5.3), and its eigen-value. For this it is convenient to rewrite our equation as follows:

$$\begin{aligned} x &= (1 + e^{-s})^{-1} , \\ \varphi(x) &= x(1-x)\psi(s) , \end{aligned} \tag{5.10}$$

so that $x = 0$ corresponds with $s = -\infty$ and $x = 1$ with $s = +\infty$.

Let us introduce the Fourier transform $\psi^F(p)$ of $\psi(s)$. We get

$$\hat{K}\psi + \hat{V}\psi = 0 , \quad \text{where}$$

$$\begin{aligned} \hat{K}\psi^F(p) &= (\pi p \coth \pi p - 1)\psi^F(p) , \\ \hat{V}\psi(s) &= \left(\frac{\alpha_1+1}{1+e^{s}} + \frac{\alpha_2+1}{1+e^{-s}} - \frac{\mu^2}{(1+e^{s})(1+e^{-s})} \right)\psi(s) , \\ \psi(s) &= \int_{-\infty}^{\infty} e^{ips}\, \psi^F(p)\, dp . \end{aligned} \tag{5.11}$$

If $\hat{V}$ is small, then $\hat{K}$ must be small, hence $\psi^F(p)$ must vanish rapidly when $p \neq 0$. So $\psi(s)$ will be a slowly varying function. We may make a Taylor expansion of K:

$$K(p) \simeq \frac{\pi^2 p^2}{3} \quad ; \quad \hat{K} \simeq -\frac{\pi^2}{3}\left(\frac{d}{ds}\right)^2 . \tag{5.12}$$

At $s \to \infty$, we have

$$\frac{\pi^2}{3}\left(\frac{d}{ds}\right)^2\psi = (\alpha_2+1)\psi = \frac{\pi}{g^2 N} m_2^2 \psi ,$$

so $\psi(s) \to e^{-\beta_2 s}$ with

$$\frac{\pi^2 \beta_2^2}{3} = \frac{\pi m_2^2}{g^2 N} . \tag{5.13}$$

Similarly, for $s \to -\infty$

$$\psi(s) \to e^{\beta_1 s}$$

By integrating the equation over s from before to after the dip in V(s), we get

$$\beta_1 + \beta_2 = \frac{3}{\pi^2} \mu^2 , \tag{5.14}$$

and so we get

$$\mu^2 = \frac{\pi^{3/2} (m_1 + m_2)}{g\sqrt{3N}} . \tag{5.15}$$

and the mass M_o of the lowest bound state satisfies

$$M_o^2 = g\sqrt{\frac{\pi N}{3}} (m_1 + m_2) . \tag{5.16}$$

We can express g in terms of α', using eq. (5.8):

$$M_o^2 = \frac{m_1 + m_2}{\sqrt{3\alpha'}} . \tag{5.17}$$

This formula holds in the limit of small m_1, m_2 and the coefficient is exact in two dimensions. Let us for a moment speculate that it also holds in four dimensions. Then

$$\begin{aligned} m_\pi^2 &= (m_p + m_n) \cdot 600 \text{ MeV} ; \\ m_p + m_n &= 30 \text{ MeV} ; \\ m_K^2 &= (m_N + m_\lambda) \cdot 600 \text{ MeV} ; \\ m_\lambda &= 400 \text{ MeV}. \end{aligned} \tag{5.18}$$

The last value of 400 MeV is probably too large. We expect (5.17) to be wrong for the λ quark. We know that for very large m_1, m_2 :

$$M_o = m_1 + m_2 \,. \tag{5.19}$$

An extrapolation between (5.17) and (5.19) is for instance

$$M_o^2 = \left(\frac{1}{\sqrt{3\alpha'}} + M_o\right)(m_1 + m_2) \,. \tag{5.20}$$

which would yield the more reasonable value

$$m_\lambda = 200 \quad MeV \tag{5.21}$$

The proton-neutron quark mass difference can be estimated by comparing $K^{\pm}$ and K^o. We get values of several MeV. A rough estimate of the quark masses, as they occur in the Lagrangian, is

$$m_p = 10 \text{ MeV}$$
$$m_\pi = 15 \text{ MeV}$$
$$m_\lambda = 200 \text{ MeV}.$$

6. COMPARISON WITH THE σ-MODEL

We would like to have a qualitative explanation of this peculiar behaviour of the lowest bound state mass, as the quark masses go to zero. If we restrict ourselves to the case that the two quarks are equal, then the explanation is the following.

The fact that there is a physical zero-mass, pseudoscalar, bound state if $m_q = 0$ indicates that chiral symmetry is spontaneously broken. Indeed, the bound states do not occur in degenerate parity doublets (one can show that the states have alternatingly positive and negative parity).

Let us therefore assume that the lowest two states, which we shall call π and σ, are roughly described by an effective chirally invariant Lagrangian:

$$\mathcal{L} = -\tfrac{1}{2}(\partial\pi)^2 - \tfrac{1}{2}(\partial\sigma)^2 - \frac{\lambda}{8}(\pi^2+\sigma^2-F^2)^2 + \Delta\mathcal{L} \ . \tag{6.1}$$

Here the pion field roughly corresponds to

$$\pi = C_\pi \bar\psi \gamma_5 \psi \ ,$$

and the sigma field

$$\sigma = C_\sigma \bar\psi \psi \ ,$$

$\Delta\mathcal{L}$ stands for the symmetry breaking term, generated by the mass m of the constituent quarks:

$$\Delta\mathcal{L} \simeq -m\bar\psi\psi = -C_\sigma m\sigma = -c\sigma \ . \tag{6.2}$$

The coefficient c is linear in m for small m. It generates a mass for the pion. The vacuum expectation value of the σ field is now

$$\langle\sigma\rangle^2 = F^2 + \frac{2c}{\lambda|\langle\sigma\rangle|} \simeq F^2 + \frac{2c}{\lambda F} \ . \tag{6.3}$$

If we expand the potential in (6.1) around the new vacuum we get

$$\begin{aligned} &-\frac{\lambda}{8}\left(\pi^2+\frac{2c}{\lambda F}\right)^2 + \cdots \\ &= -\frac{\lambda}{8}(\pi^2)^2 - \frac{c}{2F}\pi^2 + \cdots \end{aligned} \tag{6.4}$$

Clearly, the pion mass satisfies

$$m_\pi^2 = \frac{c}{F} \ , \quad \text{if } c \text{ is small.} \tag{6.5}$$

This explains why m_π^2 starts being linearly proportional to m_q.

The model with negative m_q formally exists and is equivalent to that with positive m_q through a γ^5 transformation. Therefore we should really say

$$m_\pi^2 = C^{st} |m_q| \quad . \tag{6.6}$$

Thus there is a singularity at $m_q = 0$. This is typical for theories with spontaneous symmetry breakdown. It is a consequence of the fact that $\langle\sigma\rangle$ makes a sign flip when c goes through zero.

7. THE MASS OF THE ETA PARTICLE

If the proton and the neutron quark are both light, then our model predicts four light bound states according to eq. (5.17). These would be $\pi_\pm$, π_o and η. They are the Goldstone bosons of chiral $U(2) \times U(2)$. Indeed, in four dimensions also, if $m_{p,n} \to 0$ there should be a chiral $U(2) \times U(2)$ symmetry, rather than $SU(2) \times SU(2)$. This raises the question: why is the eta not as light as the pions, or, why is the chiral subgroup $U(1)$ badly broken if the masses of the quarks are light?

In the following arguments we disregard strangeness, because extension from $SU(2)$ to $SU(3)$ is straightforward and the basic problem remains the same.

Clearly there is an anomalous phenomenon, affecting only the iso-singlet state. This suggests that we have to look at virtual pair annihilation and -recreation, or closed loops of quarks.

The resolution of the problem was suggested by Fritzsch, Gell-Mann and Leutwyler [7]: Chiral $U(1)$ symmetry is broken by a Bell-Jackiw triangle

anomaly. The isosinglet axial vector current conservation is not only broken by the quark masses, but also the gluon field:

$$\partial_\mu j_5^\mu = 2m j_5 + C \varepsilon_{\mu\nu\alpha\beta} \mathrm{Tr}\, G_{\mu\nu} G_{\alpha\beta}\,, \tag{7.1}$$

where the last term is an inevitable consequence of color-gauge invariant renormalization.

Unfortunately this does not solve the problem completely, because how can the last term shift the mass of the Goldstone particle away from zero? It should cancel the Goldstone pole in the axial current matrix elements, but it seems to vanish at $q \to 0$. Stated differently: it is easy to redefine the axial vector current in such a way that it is conserved:

$$\tilde{j}_5^\mu = j_5^\mu - 2C \varepsilon_{\mu\nu\alpha\beta} \mathrm{Tr}\, A_\nu \left(\partial_\alpha A_\beta - \partial_\beta A_\alpha + \tfrac{2}{3} g [A_\alpha, A_\beta]\right); \tag{7.2}$$

$$\partial_\mu \tilde{j}_5^\mu = 0.$$

Now several arguments have been presented in the literature to talk away this discrepency. The point is that the correction to j_5^μ is not color-gauge invariant and might therefore show unphysical poles at $q^2 = 0$, or in coordinate space, the correction might grow indefinitely as a function of distance. We may expect such things in theories that confine quarks. But how can we understand this behaviour?

It is here also that two dimensional gauge theories show their merits. In two dimensions Bell-Jackiw anomalies also exist, but in two-point functions, not triangle loops. And the anomaly in two dimensions does not occur when the color group is SU(N), but when there is a U(1) component (contrary to the four dimensional case, U(1) photons also confine quarks, in two dimensional theories).

The axial current in two dimensions satifies

$$\partial_\mu j_5^\mu = 2m j_5 + C \varepsilon_{\mu\nu} F_{\mu\nu} . \tag{7.3}$$

Here also, we find that the correction is a pure divergence: we can define

$$\tilde{j}_5^\mu = j_5^\mu - 2C \varepsilon_{\mu\nu} A_\nu , \tag{7.4}$$

so that

$$\partial_\mu \tilde{j}_5^\mu = 0. \tag{7.5}$$

Nevertheless, the anomaly raises the mass of isosinglet pseudoscalars. In the limit $m \to 0$ these are the only states that can be calculated for all color gauge groups SU(N) X U(1). In two dimensions:

$$j_5^\mu = \varepsilon_{\mu\nu} j^\nu = - \varepsilon_{\mu\nu} \partial_\alpha F_{\alpha\nu} \tag{7.6}$$

Let

$$\tilde{F} = \varepsilon_{\mu\nu} F_{\mu\nu} ;$$

$$\varepsilon_{\mu\nu} F_{\alpha\nu} = \frac{1}{2} \delta_{\mu\alpha} \tilde{F} . \tag{7.7}$$

Then we have from eq. (7.3),

$$\partial_\mu j_5^\mu = C \tilde{F} ;$$

$$\partial_\mu j_5^\mu = - \frac{1}{2} \partial^2 \tilde{F} ; \tag{7.8}$$

$$(\partial^2 + 2C) \tilde{F} = 0 .$$

We have a free particle with a mass due to the anomaly. The same can be obtained by summing all diagrams. In fact this particle is identical to the massive "photon" obtained in the Schwinger model [8]. The reason why the mass does not vanish is that the correction term in (7.4) does not vanish at large distances. We now believe that the four-dimensional

theory with its one-dimensional flux tubes, as described in the introduction, is just as badly infrared divergent, so that the correction terms in (7.2) need not vanish at large distances either.

8. THE $\eta \to 3\pi$ DECAY

Another problem often cited in connection with the eta particle is its decay to three pions. This however is not a fundamental problem in the scheme of this theory. Only if one assumes it to be a purely electromagnetic decay then current algebraic arguments show that it should be suppressed by factors m_π^2/m_η^2 and the feinstructure constant α, but in fact the decay is known to be rather strong.

However, if one introduces electromagnetism in the theory, then one should either introduce a counter term corresponding to the proton-neutron quark mass difference, or extend towards a unified theory of electromagnetic and weak interactions so that such a counter term is not arbitrary but can be calculated. In either case we have already an estimate of what the mass difference should be, and in either case the current algebra argument would be wrong. We can estimate the $\eta \to 3\pi_o$ decay rate in an effective chiral Lagrangian theory. We have effective fields $\varphi_{\bar{q}_1 q_2}$ for the $\bar{q}_1 q_2$ pseudoscalar bound states. The relevant part of the Lagrangian is

$$\mathcal{L} = -\tfrac{1}{2}(\partial\varphi_{\bar{p}p})^2 - \tfrac{1}{2}(\partial\varphi_{\bar{n}n})^2$$
$$-\tfrac{1}{2}m_\pi^2(\varphi_{\bar{p}p}^2 + \varphi_{\bar{n}n}^2) + \tfrac{1}{4}\Delta m^2(\varphi_{\bar{p}p}^2 - \varphi_{\bar{n}n}^2) \qquad (8.1)$$
$$-\tfrac{\lambda_1}{8}(\varphi_{\bar{p}p}^4 + \varphi_{\bar{n}n}^4) - \tfrac{\lambda_2}{4}\varphi_{\bar{p}p}^2\varphi_{\bar{n}n}^2 - \tfrac{1}{4}m_\eta^2(\varphi_{\bar{p}p} + \varphi_{\bar{n}n})^2 .$$

Here Δm^2 is proportional to the proton-neutron quark mass difference and is a few $(\mathrm{MeV})^2$. The last term was due to the Bell-Jackiw anomaly.

There is no reason for λ_1 and λ_2 to be equal (In the $N \to \infty$ limit $\lambda_2/\lambda_1 \to 0$, because the λ_2 term is non-planar).

Diagonalization gives

$$\varphi_{\bar{p}p} = \frac{1}{\Pi}\left(\pi_0 + (1+\varepsilon)\eta\right) ,$$
$$\varphi_{\bar{n}n} = \frac{1}{\Pi}\left(-(1+\varepsilon)\pi_0 + \eta\right) . \qquad (8.2)$$

Here Π is a normalization constant and

$$\varepsilon = {\Delta m^2}/{m_\eta^2} . \qquad (8.3)$$

In the diagonalized fields the interaction term gives, among others, a term

$$(\lambda_1 - \lambda_2)\,\varepsilon\,\pi_o^3\eta , \qquad (8.4)$$

which is of the correct order of magnitude. No symmetry principle forbids such a term.

REFERENCES

1) G. 't Hooft, talk given at the I.P.S. Conference on High Energy Physics, Palermo (Italy) 23 - 28 June 1975, and references therein.

2) G. 't Hooft and M. Veltman, "DIAGRAMMAR", CERN Report 73/9 (1973), reprinted in "Particle Interactions at Very High Energies", NATO Advanced Study Inst. Series, Sect. B, Vol. 4b, p. 177.

3) G. 't Hooft, Nucl. Phys. B72 (1974) 461.

4) G.P. Canning, Niels Bohr Inst., prepr. NBI-HE-74-2 (1974).

5) S. Mandelstam, Dual Resonance Models, Physics Reports, to be published.

6) G. 't Hooft, Nucl. Phys. B75 (1974) 461.

7) H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys. Lett. 47B (1973) 365.

8) J. Schwinger, Phys. Rev. 128 (1962) 2425.

ACKNOWLEDGEMENT

The author would like to thank D.A. Ross and M. Veltman for many interesting discussions.

D I S C U S S I O N

CHAIRMAN: G.'t Hooft

Scientific Secretaries: M. Peskin, R.C. Shellard

DISCUSSION No. 1

- *PETRONZIO:*

Does your 1/N expansion incorporate the Zweig rule?

- *'t HOOFT:*

On the first level, the 1/N expansion obviously implies Zweig's rule, which is just the suppression of decays which cannot procede straightforwardly through planar duality diagrams. In practice the problem is more puzzling: we get a suppression factor of 1/N for a nonplanar contribution, but N is only 3 in the real world. In the case of charmonium, I think you also get a dynamical suppression due to the asymptotic freedom of the theory. David Gross will tell more about this in his lectures.

- *SHELLARD:*

A graph with a fermion loop has many more configurations than just a planar graph, graphs with worm holes even more. Would that not overcome the 1/N and $1/N^2$ factors, when you sum over all possible configurations?

- *'t HOOFT:*

I think the essential properties of a certain order in $\tilde{g}$ are already contained in the planar graphs. The non-planar graphs in the form of dual loops will represent unitarity corrections to the meson-meson scattering amplitudes. The factor 1/N will be exactly like a free coupling constant that multiplies these corrections.

- *ROTHE:*

This morning you expressed your belief that the study of a gauge theory in 2-dimensions would exhibit many interesting features present in a non-abelian gauge theory in 4 dimensions. But to my mind one of the very interesting new ingredients coming into such a non-abelian gauge theory is the presence of higher order self-couplings of the gauge fields which are expected to be crucial for explaining quark confinement. Such coupling terms are absent in a 2-dimensional theory. Are you not loosing thereby one of the most interesting features of non-abelian gauge theories?

– *'t HOOFT:*

I claimed that, in 4-dimensions, these non-linear couplings distort the gauge field of two charges into a narrow tube. In two dimensions this sort of behaviour arises in a trivial way: the whole world is aline and the gauge field must lie only along this line. In 2-dimensions, then, one easily finds a linearly rising potential. Our expectations for 4-dimensions are that the potential between the quarks is Coulombic at small separations but very quickly begins to show the same linear rise. The 2-dimensional model gives no explanation of this behaviour of the four dimensional theory, but, rather contains this result in a trivial way. I should add that the 2-dimensional model contains other very interesting features, in particular, the fact that SU(2) × SU(2) is realized in the Goldstone mode. This may mean that the 2-dimensional theory is already quite close to 4-dimensions, where we also see that SU(2) × SU(2) is realized in the Goldstone mode.

– *COLEMAN:*

I was under the impression that spontaneous breakdown of continuous symmetries was impossible in 2-dimensions.

– *'t HOOFT:*

I find the Goldstone phenomenon only in the limit $N \to \infty$. In this limit the Goldstone bosons are non-interacting. A non-interacting theory is the one case in which your theorem is not applicable.

– *CANNING:*

What do the baryon graphs look like in your theory? They do not seem to be restricted to a definite topology.

– *'t HOOFT:*

To obtain baryons straightforwardly, one must really be working with SU(3). In the large N limit there is a way to make baryons, it is not a very beautiful way, but it can be done. Define the P and N quarks to carry lower indices: P_i, n_i and the λ quark to carry two upper colour indices λ^{ij}, so that $\lambda^{ij} = -\lambda^{ji}$. If we are working in SU(3), we can write $\lambda^{ij} = \varepsilon^{ijk}\lambda_k$. This takes us back to the normal definition of a λ quark. In SU(N) there is no simple connection between upper and lower indices and the λ is no longer coloured like an ordinary quark. In SU(3) we are not restricted to planar graphs. But in the large N limit the only graphs that survive have the form of Fig. 1; that is planar graphs bounded by the P and N quarks, with the double-edged λ line propagating through the centre

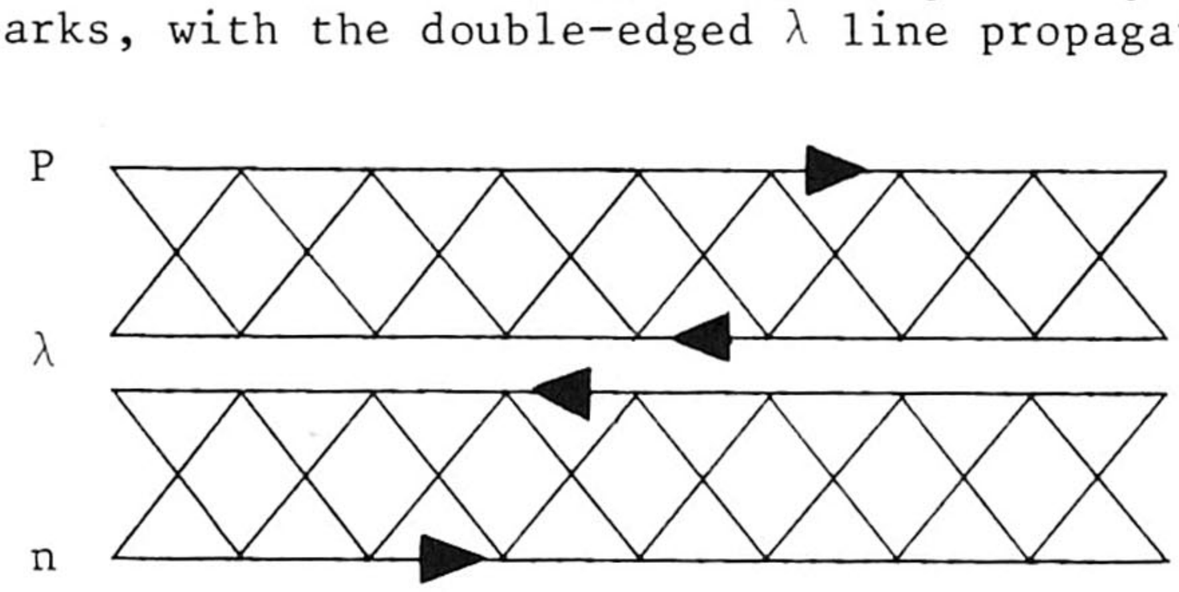


Fig. 1

- *MILLER:*

Well, what is an Ω^- then?

- *'t HOOFT:*

That is a good question. You cannot get an Ω^- this way.

- *JACOBS:*

Do you know whether your can consistently calculate the corrections in 1/N in your expansion? There are some field theories in which one can see the 1/N expansion breaking down in the first few orders.

- *'t HOOFT:*

Yes, in scalar field theories the 1/N expansion shows spurious tachyons poles in the first few orders in 1/N, coming from graphs that have a chain of bubbles. But I think this behaviour is related to the fact that if one sets up the Callan-Symanzik equation in these theories, one sees a pole in the effective coupling constant. This is a strange and quite unphysical divergence; such a divergence does not appear to occur in asymptotically free theories. Hence, we can hope that asymptotically free theories avoid these difficulties though we do not really know.

- *FRAMPTON:*

In the $N \to \infty$ limit, you showed that only the planar diagrams survive. This does not seem to correspond really to non-interacting mesons.

- *'t HOOFT:*

My point is that the diagrams of the form of Fig. 2, are down by a factor 1/N, relative to the corresponding diagram without the internal quark loop, so the corresponding dual coupling constant is vanishing. In dual theories the diagram with that topology carries as an extra factor the dual coupling constant.

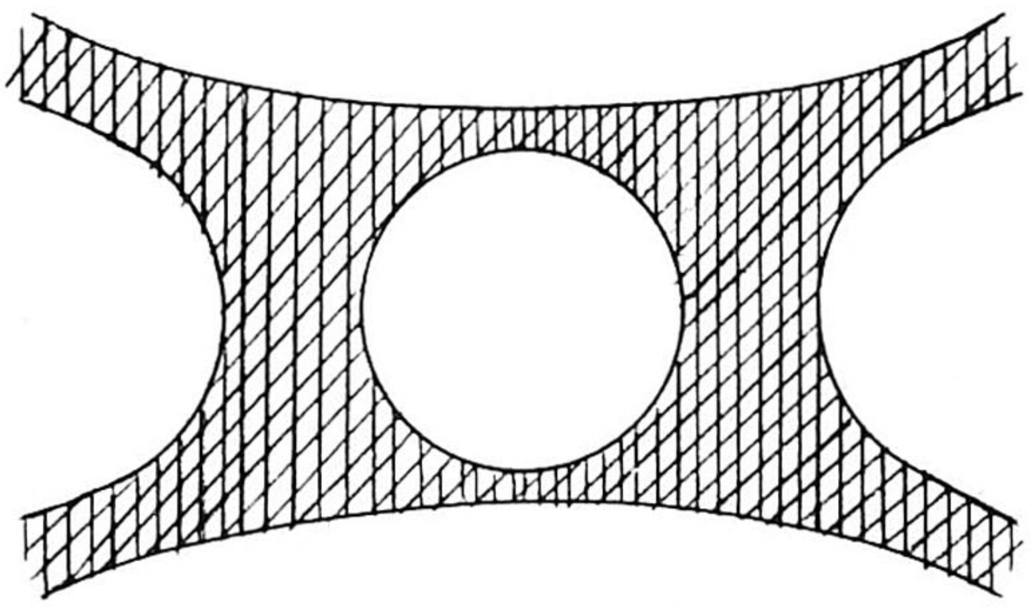


Fig. 2

- *FRAMPTON:*

That means that you are still left with the tree amplitudes, does it not?

- *'t HOOFT:*

Since the dual coupling constant tends to zero the imaginary part of the scattering amplitude vanishes and the mesons do not interact. I think this corresponds to the fact that the total widths are zero in the born amplitude.

- *FRAMPTON:*

I am not sure it is accurate to describe the tree amplitude as really non-interacting, but that can be just a semantic question.

- *'t HOOFT:*

The poles are on the real axis, the corrections for unitarity tend to zero and the parameters characterizing the dual loop expansion are zero. That the born amplitude do not seem to vanish is only due to the overall normalization.

DISCUSSION No. 2 *(Scientific Secretaries: R.C. Shellard, M.E. Peskin)*

- *MENOTTI:*

Gauge theories are troubled by infrared divergences; one hopes that these infrared problems will go away in the exact solution to the theory. Your 2-dimensional model provides an example of this. But are there any approximation schemes for explicitly summing the infrared divergences in a four-dimensional gauge theory, perhaps by summing selected classes of graphs?

- *'t HOOFT:*

I know of no way to see explicitly that the infrared divergences cancel. You can write down models representing what you expect to happen. For example, add to the normal gauge field action ($-\frac{1}{4}G_{\mu\nu}G^{\mu\nu}$) a scalar field ϕ, in the following way:

$$\mathcal{L} = -\tfrac{1}{4}Z(\phi)G_{\mu\nu}G^{\mu\nu} + \tfrac{1}{2}(\partial\mu\phi)^2 - \tfrac{1}{2}m^2\phi^2$$

where $Z(\phi)$ is a function of ϕ which becomes zero as ϕ goes to zero. Then, if the field goes to zero at infinity, the gauge field exists only in a finite region. (The vacuum, of course, has a very crazy appearance, because it contains no gauge field degrees of freedom.) If you believe that dynamical effects might produce such a situation, you can consider this to be a model of the real world -- it is a sort of effective bag model. But this, really, is the best I can do. I know of no successful attempt to sum the infrared divergences in a real Yang-Mills theory.

- *CALDI:*

You stated that you have a Goldstone boson in your 2-dimensional model in the limit $N \to \infty$, because in that limit there are no interactions and so Coleman's theorem does not hold. However, as far as I have understood it, to get a Goldstone boson it is essential to have interactions since, after all, the Goldstone boson is supposed to be some sort of collective excitation. So precisely how is your zero mass state a Goldstone boson?

- *'t HOOFT:*

I must make my statement more precise. If you compare this to the σ model:

$$\mathcal{L} = \mathcal{L}_{\text{kinetic}} - V(\phi) \ ,$$

with $V(\phi)$ having its minima at $\phi \neq 0$, you see it is the interactions in lowest order which give the Goldstone state. In my model, as $N \to \infty$, it is the interactions among the quanta of fields that disappear, not the interactions of the fields themselves. In other words, at the limit $N \to \infty$, we may neglect quantum effects. Coleman showed that it is the loop corrections (i.e. quantum effects) which destroy the Goldstone mode in 2-dimensions, due to infrared divergences. It is these that are neglected in the $N \to \infty$ limit, since they are at most of order $1/N$. So there are interactions, but you may neglect the troublesome higher-order ones.

- *COLEMAN:*

It is only this formal limit that displays the Goldstone phenomenon. There is no consistent Goldstone phenomenon in 2-dimensional quantum field theory.

- *'t HOOFT:*

I agree. The whole point is that I take first the limit $N \to \infty$, then $m \to 0$. If the Goldstone boson were also to turn up when $m \to 0$ with N finite, then that would contradict what Coleman says.

- *GROSS:*

I should note that your expansion parameter is really $\left\{(1/N)\cdot\left[\log(m^2/g^2N)\right]\right\}$. The higher corrections will be log divergent as $m \to 0$.

- *FRAMPTON:*

Let me ask about your linear or approximately linear trajectories. In one space plus one time dimension there is no angular momentum. Thus what you are looking at are the pion and its radial recurrences. Now, phenomenologically, we know nothing about that. My question is, then: do you think this result has anything to do any experimental observations, such as the fact that the parent trajectories are approximately linear?

– *'t HOOFT:*

It is best to think of the dual resonance model! In three space dimensions the flux lines should make a flux tube or dual string. The parent trajectories should be linear with parallel daughters, and then one has equally-spaced spin zero $\pi - \sigma - \pi - \sigma - \ldots$ states.

– *FRAMPTON:*

So you regard this as the spin-zero line of the usual Chew-Frautschi plot?

– *'t HOOFT:*

Yes.

– *PESKIN:*

Conventionally, in looking for chiral symmetry breaking, one assumes the existence of a fermion mass, or a tadpole, or some other indication of an alteration in vacuum structure; then one solves for it self-consistently. How did you introduce such new structure of the vacuum into this calculation?

– *'t HOOFT:*

One needs self-consistent assumptions to break chiral symmetry only if one's theory is chirally symmetric. In my calculation one starts with massive quarks, so that chiral symmetry is already broken; the limit $m \to 0$ is taken only after the dynamical problem has been solved. We find, as we should, the phenomena associated with broken chiral symmetry. But the method of calculation makes it very difficult to explore the properties of the broken-symmetry vacuum.

– *PETRONZIO:*

You presented two different expressions for the mass of the bound state of two quarks in terms of the masses of the two quarks: one which holds for small quark masses, and another which holds for large quark masses. What sets the scale of the masses of the quarks?

– *'t HOOFT:*

In two dimensions, the coupling constant g is dimensional and gives you a natural scale for the quark mass: $g/\sqrt{\pi}$.

– *PETRONZIO:*

But, in this case, what is the meaning of your extension to four dimensions, where the coupling parameter g is dimensionless?

– *'t HOOFT:*

In general, the dimension D(g) of the coupling constant g is a function of the number of space-time dimensions d. The relation, in units of mass, is the following

$$D(g) = \frac{4 - d}{2} \quad .$$

One can renormalize the theory in 4-dimensions by taking the limit $\varepsilon \to 0$ of the same theory in $4 - \varepsilon$ dimensions. In this procedure one has a coupling constant of dimension $\varepsilon/2$, so a natural scale for the quark masses is still present: it is proportional to $(g)^{2/\varepsilon}$.

- *POLITZER:*

An alternative description (in the context of asymptotic freedom) is that in 4-dimensions the renormalization scale needed to define the coupling constant provides a scale for measuring quark masses. For instance, if the coupling is large when defined on the scale of a quark mass, then that quark is light, and the relativistic formula for bound state masses is plausible as a limit. If, however, the coupling is small on the scale of some other quark mass, that quark can be regarded as heavy because small coupling suggests small "binding energies" and hence a non-relativistic mass formula.