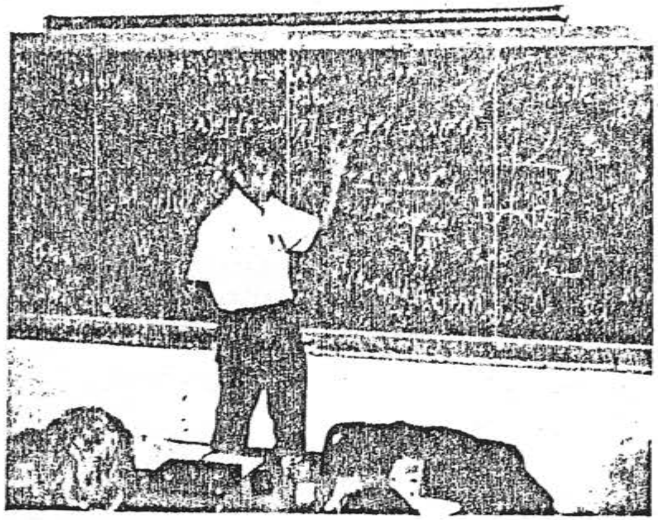


COURSE 2

# GAUGE THEORIES

Richard P. FEYNMAN

*California Institute of Technology,*
*Pasadena, Calif. 91100, USA*

*Lecture notes by K.H. Craig and I. Hinchliffe*

## Contents




*R. Balian and C.H. Llewellyn Smith, eds., Les Houches, Session XXIX, 1976 – Interactions électromagnétiques et faibles à haute énergie/Weak and electromagnetic interactions at high energy*




## 1. Introduction

In 1954 Yang and Mills [1] wrote a paper in which they made a theory in analogy with QED for a system in which a particle could carry more than one "charge". In those days there were particles like protons and neutrons which were thought to be one object – a nucleon which appears in two guises – proton and neutron. This is the theory of isospin in which the analogue of charge is $I_z$, the $z$-component of isospin; $+\frac{1}{2}$ for proton, $-\frac{1}{2}$ for neutron. Then there was the question of what field, analogous to the photon in QED, could interact with such a charge. In the case of QED the photon is a vector field; therefore Yang and Mills tried to make a theory of a vector field interacting with "charges" that might have more than one value. There had already existed for some time a theory of (pseudo) scalar fields which could interact with particles of different "charge" but their properties weren't as interesting as those of vector fields. Since the "charge" can be flipped back and forth, the fields which are coupled to the "charges" are more interesting than that of QED. Finally the theory had very great beauty and simplicity.

We human beings see in a symmetrical theory a certain beauty; the Greeks, for example, saw in the theory of the planets that they went around in circles at a uniform speed, a phenomenon which today, we would characterise by a group theoretic property: the orbit is such that a displacement in time is equivalent to a rotation.

Today we still have this desire to see symmetrical things and therefore the Y–M theory looks very good. In contrast to QED, here we have a field with many components which couple to different "charges". This is because the field, in addition to the neutral component which couples to + and – charge, also has charged components which flip one "charge" into another. Therefore in the case of isospin we have a source of isospin $\frac{1}{2}$ (the nucleon) and a field of isospin 1. Now this theory was beautifully symmetric but it did not agree with experiment; although isospin is almost exactly conserved the theory is similar to electrodynamics in that the mass of the vector field is zero, but there is no obvious long range force between nucleons: so it's wrong. The first hope was to put a mass in somehow, but that destroyed the symmetry and consequently the beauty. What were the vector particles anyhow? They were supposed to be $\rho$ mesons, but there seemed to be nothing more fundamental about the $\rho$ meson than all the other hadrons. So the idea that one of these hadrons was the fundamental field was lost.

Later people such as Goldstone [2] began to look at broken symmetries; Higgs [3,4] and Kibble [5] found that a massless vector theory with broken symmetry was in some sense equivalent to one with mass. Today the only symmetries we see in nature are isospin (a near perfect symmetry) and SU(3) (a clear, but imperfect symmetry) and we would, therefore, think that, because we like symmetries so much, the excitement of the day would be that we had an understanding of these symmetries at last. We do not. In fact, the Y–M theory with broken symmetry is assumed to apply somewhere else.

In the meantime, there was developed a weak interaction theory in which, in one interpretation, one had vector mesons with mass. Taken directly, a field theory of massive vector mesons is highly non-renormalisable. Such a theory works fine as long as we only work with first order diagrams. However, attempts to go to higher order lead to unremovable divergences. It is necessary to go to higher order for two reasons:

(1) It is not sensible to have a theory which only works to first order.

(2) Consider a process where the amplitude is calculated to first order in $g$. The probability of the process occuring is $O(g^2)$. Therefore the probability of non-occurrence is $\sim 1 - O(g^2)$. Hence the amplitude for non-occurrence $\sim \sqrt{1 - O(g^2)} \sim 1 - O(g^2)$. Therefore one must know something about amplitudes to order $g^2$. It is, therefore impossible to have a theory which only works to first order if one wants to conserve probability. People did not worry about this until recently. When they did, they found that if they started with one of these symmetric theories and used the Higgs mechanism to add mass they might be able to represent these vector mesons (intermediate vector bosons) with mass in a way that was renormalisable in the same sense as QED. This was subsequently proved in detail, and the consequences of these theories are, therefore, as calculable as those of QED.

Sunsequently it was found by Llewellyn Smith [6], that if one starts with massive vector mesons and requires renormalisability one is driven to Y–M theories with broken symmetries, provided one is prepared to introduce new particles to cancel divergences. One has a choice of such particles and a particular choice leads to a particular model. It is important to realise that there is no unique prescription for doing this. One can also use different symmetry groups and different methods of breaking the symmetry. It is therefore, not true to say that these theories make an unambiguous prediction of the existence of neutral currents, because one can always take a different theory which has no neutral current but some other new particle(s) (e.g. "heavy" lepton). It appears that there is now experimental evidence for both neutral currents and heavy leptons.



In hadron physics the quark theory was evolved and found to be paradoxical. Protons are supposed to be made out of three quarks as is the $\Delta^{++}$, which has spin $\frac{3}{2}$. Consider the case where $J_z = +\frac{3}{2}$:

$$\begin{matrix} u & u & u \\ \uparrow & \uparrow & \uparrow \end{matrix} .$$

The dynamical theory says that the quarks are in a relative S-state in order to get the right order of magnitude for matrix elements and magnetic moments; then we have three particles in the same state. There is a problem in that it had been proved that we can't put three spin $\frac{1}{2}$ particles in the same state. This proof assumed that the particles could be separated from each other. We also know that quarks don't seem to appear as free particles. It is, therefore, not clear that the proof holds for quarks. Nevertheless, to be conservative, we will accept the theorem, and so a simple explanation of the problem is that the three quarks are different, i.e. we assign them a new quantum number (colour) which takes three values A, B, C:

$$\begin{matrix} u_A & u_B & u_C \\ \uparrow & \uparrow & \uparrow \end{matrix} .$$

There are, at present, only two places in the experimental world where the colour hypothesis can be checked:

(1) A subtle test is connected with the anomaly in the $\pi^0 \to 2\gamma$ decay. It turns out that a theory without colour gives a decay rate a factor of three too small; a theory with three colours agrees with experiment.

(2) The ratio

$$R = \frac{\sigma(e\bar{e} \to \text{hadrons})}{\sigma(e\bar{e} \to \mu\bar{\mu})} ,$$

is equal to the sum of the squares of the quark charges and equals 2/3 for u, d, s without colour. In fact, the data shown symbolically* in fig. 1.1 are in disagreement with this value. u, d, s with colour give $R = 2$ in better agreement below 4 GeV.

Another possible experimental test of colour is lepton production in pp collisions by the Drell–Yan mechanism but the evidence is inconclusive. The amount of experimental evidence for the beautiful symmetry of colour is not

* See the lectures by G. Wolff and B.H. Wiik for a discussion of the data.

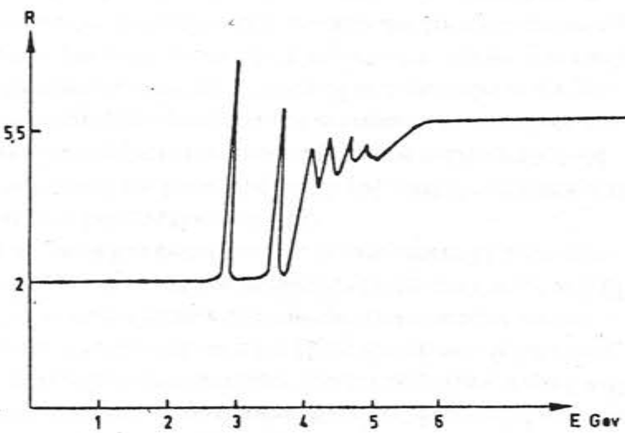



Fig. 1.1.

great but is theoretically very strong because it helps to explain a number of other features e.e. why three quarks in a baryon held together.

This theory of colour is so symmetrical that a good guess is that hadrons are made of quarks with flavours u, d, s, c, ... and three colours A, B, C with exact colour symmetry. It turns out that the colour couples to an eight component field. The reason why there are eight components is as follows:

When a field quantum (gluon) is emitted, the quark colour may or may not change. There are nine ways of coupling a gluon between an initial (three colour possibilities) and a final (three colour possibilities) quark.

| | |
|---|---|
| A → B | |
| A → C | A → A |
| B → A | B → B |
| B → C | C → C . |
| C → A | |
| C → B | |

But the linear combination of the components of the vector field which couples equally to all the quarks (the singlet) need not be in the theory, all its properties are independent of the other eight components. Under a linear transformation of the colours, the eight mix together so all are necessary, but the singlet stays unchanged. Hence we are left with eight components. (Whether we add the ninth or not, and with what coupling, is up to us, but we will leave



it out as apparently unnecessary at present.) This theory with exact SU(3) colour symmetry is called quantum chromodynamics (QCD).

At first sight we have a problem, viz., massless vector meson's would imply a long range force between quarks and so we could separate them. Experimentally this does not seem to be true. One solution to the problem is that QCD is wrong. Another way out is to say that we do not understand the consequences of Y–M well enough and that at large distances the forces might become large enough to confine the quarks. That is the foremost problem of QCD. Also there are infrared divergences in QCD which are more serious that in QED and the method to handle them is not yet known. In spontaneously broken Y–M theories these infrared problems are absent.

To summarize, there are two applications of Y–M theories: with broken symmetry in weak interaction (e.g. Weinberg [7] – Salam [8] model) and in strong interactions by QCD with perfect unbroken colour symmetry.

## 2. Classical Yang–Mills theory

We are going to take a more or less elementary and direct view of Yang–Mills theory, rather like the authors did. We start with the example of SU(2) in which we have a 2 component spinor representing proton and neutron. We start with the Lagrangian density for the free proton and neutron fields

$$\mathcal{L}_F = i\bar{\psi}_p \partial\!\!\!/ \psi_p + i\bar{\psi}_n \partial\!\!\!/ \psi_n - m_p \bar{\psi}_p \psi_p - m_n \bar{\psi}_n \psi_n \,, \tag{2.1}$$

where we use the notation of Bjorken and Drell [9].

In the old days people wanted to add interactions with pseudoscalar particles. Consider 3 pseudoscalar particles (e.g. the pion) with charges $0(\phi_0)$, $+1(\phi_+)$, $-1(\phi_-)$; the simplest coupling to the nucleons is

$$\mathcal{L}_I = i\{\alpha\bar{\psi}_p\phi_0\gamma_5\psi_p + \beta\bar{\psi}_n\phi_0\gamma_5\psi_n + \gamma\bar{\psi}_p\phi_+\gamma_5\psi_n + \gamma\bar{\psi}_n\phi_-\gamma_5\psi_p\}. \tag{2.2}$$

If the nucleon–nucleon force is independent of the nucleon type (and $\gamma \neq 0$), then.

$$\alpha = -\beta = \gamma/\sqrt{2}\,. \tag{2.3}$$

A neater way to write the Lagrangian in this case is to introduce a spinor $\psi_a$ ($a$ = p, n) and an isovector $\phi_i$ ($i$ = 1, 2, 3) in isospin space, where

$$\phi_1 = \frac{1}{\sqrt{2}}(\phi_+ + \phi_-), \qquad \phi_2 = \frac{i}{\sqrt{2}}(\phi_+ - \phi_-), \qquad \phi_3 = \phi_0 \tag{2.4}$$

If $m_p = m_n$ the Lagrangian $\mathcal{L}_F + \mathcal{L}_I$ must be invariant under a rotation of the $\psi$ field in isospin space, i.e.

$$\psi_p \to \cos\theta\psi_p + \sin\theta\psi_n, \qquad \psi_n \to -\sin\theta\psi_p + \cos\theta\psi_n\,, \tag{2.5}$$

provided we also rotate the $\phi$ field at the same time. The Lagrangian becomes

$$\mathcal{L} = \mathrm{i}\bar{\psi}\partial\psi - m\bar{\psi}\psi + \alpha\bar{\psi}\boldsymbol{\phi}\cdot\boldsymbol{\tau}\gamma_5\psi + \text{K.E. terms for } \phi\,, \tag{2.6}$$

where

$$2\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad 2\tau_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad 2\tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.7}$$

are the Pauli spin matrices.

This is the famous pseudoscalar meson theory which was supposed to be the explanation of everything. When there were summer schools, professors explained that this was the key to the whole of strong interaction theory; they were going to explain scattering and everything else and it was just a matter of calculating the next order on a machine. But that failed so they keep on trying! The question now is, can we write a similar theory for a vector particle interacting with nucleons. It is easy to guess that a vector particle $A_\mu$ could be coupled the same way i.e.

$$\mathcal{L} = \mathrm{i}\bar{\psi}\partial\psi - m\bar{\psi}\psi + \bar{\psi}(A_\mu\cdot\boldsymbol{\tau})\gamma_\mu\psi\,, \tag{2.8}$$

where $A$ has been rescaled to absorb the coupling constant $g$. The $\gamma_\mu$ is present to contract with the space time index on the $A_\mu$ and the $\boldsymbol{\tau}$ is present to contract with the isospin index on $A$. We could also add derivative couplings on the form

$$\bar{\psi}(\partial_\mu\boldsymbol{\phi})\cdot\boldsymbol{\tau}\gamma_5\gamma_\mu\psi\,, \tag{2.9}$$

in the pseudoscalar case, and similarly in the vector case. However for simplicity we exclude these and other more complicated couplings. (There is no way a priori to exclude these couplings but theoretically there may be problems with renormalisability but in the old pion theory of nuclear forces they were all tried.)

Now the problem of coupling the vector field is very easy but it is only part of the real problem which is – what are the equations for the propagation of the field? Yang and Mills treated this problem in analogy with electrodynamics. In electrodynamics, the piece of the Lagrangian connected with the propagator of the vector fields is

$$F_{\mu\nu}F_{\mu\nu}\,,$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\,. \tag{2.10}$$



The piece of the Lagrangian analogous to (2.8) is

$$\mathcal{L} = \bar{\psi}(\mathrm{i}\partial - m)\psi + \bar{\psi}A_\mu\gamma_\mu\psi\,. \tag{2.11}$$

Why do we have such a funny looking business? Why not have, for example,

$$(\partial_\nu A_\mu)(\partial_\nu A_\mu)\,. \tag{2.12}$$

Electrodynamics has a property of gauge invariance; this means firstly if $\psi \to \mathrm{e}^{-\mathrm{i}\alpha}\psi$ and $\alpha$ is a constant, nothing happens to (2.11). But now suppose that the phase of the wave function is changed by different amounts at different space-time points i.e. $\alpha$ is a function of $x$. This is usually called a local gauge transformation. Now the kinetic energy term in (2.11) will change since

$$\mathrm{i}\bar{\psi}\partial_\mu\psi \to \mathrm{i}\bar{\psi}\partial_\mu\psi + \bar{\psi}(\partial_\mu\alpha)\psi\,. \tag{2.13}$$

One can easily make (2.11) gauge invariant by supposing that at the same time we change

$$A_\mu \to A_\mu - \partial_\mu\alpha\,. \tag{2.14}$$

Now if the theory is to be gauge invariant we cannot use (2.12) in the Lagrangian as it changes under the gauge transformation; however $F_{\mu\nu}$ is invariant so that $F_{\mu\nu}F_{\mu\nu}$ is a possible invariant contribution to the Lagrangian.

Now we can use the same trick to try to find what kind of invariant we get in this new theory with the multiple component field (isovector). Consider the transformation

$$\psi \to \exp\{-\mathrm{i}(\boldsymbol{\alpha}\cdot\boldsymbol{\tau})\}\psi\,, \tag{2.15}$$

applied to (2.8). The transformation has to be unitary and consequently it can be written in the form (2.15). We only consider the transformation to first order in $\alpha$, viz.

$$\psi \to (1 - \mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\tau})\psi\,. \tag{2.16}$$

(It can be shown that if the theory is invariant under infinitesimal transformations then it is also invariant under a finite transformation, (2.15) since this can be built up from infinitesimal transformations.) Eq. (2.8) is invariant if $\alpha$ is not a function of space-time provided

$$A_\mu\cdot\boldsymbol{\tau} \to \exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\tau})(A_\mu\cdot\boldsymbol{\tau})\exp(\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\tau})$$

i.e.

$$A'_\mu\cdot\boldsymbol{\tau} = A_\mu\cdot\boldsymbol{\tau} - \mathrm{i}[\boldsymbol{\alpha}\cdot\boldsymbol{\tau}, A_\mu\cdot\boldsymbol{\tau}] = A_\mu\cdot\boldsymbol{\tau} + (\boldsymbol{\alpha}\times A_\mu)\cdot\boldsymbol{\tau}$$

i.e.

$$A'_\mu = A_\mu + (\boldsymbol{\alpha} \times A_\mu)\,, \tag{2.17}$$

which is an infinitesimal rotation of the vector $A_\mu$ in isospace.

So far we have considered only SU(2). In another group one should have a similar sort of thing except that the $\tau$'s would be a different set of matrices, with another set of commutation relations. In general

$$[\tau_i, \tau_j] = \mathrm{i} f_{ijk} \tau_k\,, \qquad \text{(e.g. for SU(2)}, f_{ijk} = \epsilon_{ijk})\,. \tag{2.18}$$

We can generalise the notion of a vector to any group $a_i = a$ where $c = a \times b$ means $c_i = f_{ijk} a_j b_k$.

Then all the equations we write are completely general, they apply not just for SU(2). Some properties of the cross product are

$$a \cdot (b \times c) = (a \times b) \cdot c, \qquad a \times a = 0, \qquad a \times b = -b \times a,$$

$$a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0 \text{ (Jacobi identity)} \tag{2.19}$$

but note $a \times (b \times c) = b(a \cdot c) - c(a \cdot b)$ holds only for SU(2). For any group SU($n$), the number of components of $A_\mu$ is $n^2 - 1$ because the transformation parameter, $\boldsymbol{\alpha}$ can have $n^2 - 1$ components (this being the dimension of the group).

Now consider $\boldsymbol{\alpha}$ in (2.16) to be a function of space-time. The idea is that we should be able to change the phase of the wave function arbitrarily at each space-time point and transform the vector field such that the physics does not depend on this choice. We must choose the transformation of $A_\mu$ so that (2.8) is invariant under the transformation (2.16). $\boldsymbol{\alpha}$ is called the gauge parameter. Therefore under

$$\psi \to (1 - \mathrm{i}\boldsymbol{\alpha}(x) \cdot \boldsymbol{\tau})\psi, \qquad \mathrm{i}\bar{\psi}\gamma_\mu \partial_\mu \psi \to \mathrm{i}\bar{\psi}\gamma_\mu \partial_\mu \psi + \bar{\psi}\gamma_\mu (\partial_\mu \boldsymbol{\alpha}) \cdot \boldsymbol{\tau}\, \psi\,. \tag{2.20}$$

So $A$ must transform like

$$A_\mu \to A_\mu + \boldsymbol{\alpha} \times A_\mu - \partial_\mu \boldsymbol{\alpha}\,. \tag{2.21}$$

The next problem is to find a Lagrangian term for $A_\mu$ which is invariant when $A_\mu$ is changed in this manner. There are very beautiful and elegant ways of getting these things these days; but suppose that you were inventing it, what would you do to find an invariant form? You fiddle around. All the elegant stuff is found later; the way to learn is not to learn elegant things, it's to fiddle around blind and stupid. Later you see how it works; polish it up; remove the scaffolding and publish the result for other students to be amazed at your ingenuity.

For the moment forget the $\boldsymbol{\alpha} \times A_\mu$ term in (2.21), and try to find some expression like the square of the electromagnetic field tensor, i.e.



$$(\partial_\mu A_\nu - \partial_\nu A_\mu)^2\,.$$

This transforms under (2.21) as

$$\partial_\mu A'_\nu - \partial_\nu A'_\mu = \partial_\mu A_\nu - \partial_\nu A_\mu + \boldsymbol{\alpha} \times (\partial_\mu A_\nu - \partial_\nu A_\mu)$$
$$+ (\partial_\mu \boldsymbol{\alpha}) \times A_\nu - (\partial_\nu \boldsymbol{\alpha}) \times A_\mu\,. \tag{2.22}$$

If the last 2 terms were absent we would be O.K. because then

$$\partial_\mu A_\nu - \partial_\nu A_\mu\,,$$

would transform as an isovector (i.e. $V \to V + \boldsymbol{\alpha} \times V$) and therefore its square would be invariant. So we must try to get rid of the last 2 terms. Notice that the gradient of $\boldsymbol{\alpha}$ is coming from the transformation of $A_\mu$ (2.21) so that if we had a term like $A_\mu \times A_\nu$ then when we transformed it we would pick up a $A_\nu \times \partial_\mu \boldsymbol{\alpha}$. Try

$$A'_\mu \times A'_\nu = A_\mu \times A_\nu + A_\mu \times (\boldsymbol{\alpha} \times A_\nu) - A_\mu \times \partial_\nu \boldsymbol{\alpha}$$
$$+ (\boldsymbol{\alpha} \times A_\mu) \times A_\nu - (\partial_\mu \boldsymbol{\alpha}) \times A_\nu\,.$$

But

$$(\boldsymbol{\alpha} \times A_\mu) \times A_\nu = A_\nu \times (A_\mu \times \boldsymbol{\alpha})\,.$$

Add and subtract $\boldsymbol{\alpha} \times (A_\nu \times A_\mu)$ and use the Jacobi identity (eq. (2.19)) to get

$$A'_\mu \times A'_\nu = A_\mu \times A_\nu - (\partial_\mu \boldsymbol{\alpha}) \times A_\nu - A_\mu \times (\partial_\nu \boldsymbol{\alpha}) + \boldsymbol{\alpha} \times (A_\mu \times A_\nu)\,. \tag{2.23}$$

We can now get rid of the debris between (2.22) and (2.23) by defining

$$E_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + A_\mu \times A_\nu\,,$$

and so

$$E_{\mu\nu} \to E_{\mu\nu} + \boldsymbol{\alpha} \times E_{\mu\nu}\,. \tag{2.24}$$

Hence we can make an invariant quantity $E_{\mu\nu} \cdot E_{\mu\nu}$. Therefore we may write the Lagrangian density (in analogy with QED) as

$$\mathcal{L} = -\frac{1}{4g^2} E_{\mu\nu} \cdot E_{\mu\nu} + \mathrm{i}\bar{\psi}\partial\!\!\!/ \psi - m\bar{\psi}\psi + \bar{\psi}(A_\mu \cdot \boldsymbol{\tau})\gamma_\mu \psi\,. \tag{2.25}$$

Note if we rescale $A_\mu \to gA_\mu$, $\mathcal{L}$ becomes

$$\mathcal{L} = -\tfrac{1}{4} E_{\mu\nu} \cdot E_{\mu\nu} + \mathrm{i}\bar{\psi}\partial\!\!\!/ \psi - m\bar{\psi}\psi + g\bar{\psi}(A_\mu \cdot \boldsymbol{\tau})\gamma_\mu \psi\,,$$

where $E_{\mu\nu}$ is now

$$E_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + g A_\mu \times A_\nu \, .$$

This is the form given by Abers and Lee [10].

You may ask why use $E_{\mu\nu} E_{\mu\nu}$ in the Lagrangian instead of another invariant piece? For example $\widetilde{E}_{\mu\nu}$ called the dual of $E_{\mu\nu}$, defined by $\widetilde{E}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} E_{\rho\sigma}$ is such that the quantities $\widetilde{E} \cdot E$ and $\widetilde{E} \cdot \widetilde{E}$ are also invariant. The former is a pseudoscalar and one could consider it as a possible additional term in $\mathcal{L}$. It can be shown that this term has no consequences for the equations of motion provided that when $A$ is varied to obtain them one assumes that as usual there is no variation of $A$ at $\infty$. Possible consequences of a violation of this condition will be discussed later.

Now consider the action $S$ derived from the Lagrangian (2.25)

$$S = \int \{ -\frac{1}{4g^2} E_{\mu\nu} \cdot E_{\mu\nu} + \bar{\psi}(i\partial\!\!\!/ + A \cdot \tau)\psi - \bar{\psi} m \psi \} \, d^4x \, . \tag{2.26}$$

We find the equations of motion by varying the action. Varying with respect to $\bar{\psi}$ gives

$$(i\partial\!\!\!/ + A \cdot \tau)\psi = m\psi \, , \tag{2.27}$$

and a similar equation for $\bar{\psi}$ is obtained by varying with respect to $\psi$. Varying with respect to $A$ gives

$$-\frac{1}{g^2}(\partial_\mu E_{\mu\nu} + A_\mu \times E_{\mu\nu}) = J_\nu \, , \tag{2.28}$$

where

$$J_\mu = \bar{\psi} \gamma_\mu \tau \cdot \psi \, . \tag{2.29}$$

Any vector will transform as

$$\phi \to \phi + \alpha \times \phi \, . \tag{2.30}$$

The derivative of a vector transforms in a more complicated way:

$$\partial_\mu \phi \to \partial_\mu \phi + \alpha \times \partial_\mu \phi + \partial_\mu \alpha \times \phi \, . \tag{2.31}$$

Therefore if $\phi$ is a vector, its derivative is not. However we notice that

$$\partial_\mu \phi + A \times \phi \, , \tag{2.32}$$

transforms as a vector. Hence we define a covariant derivative

$$D_\mu \equiv (\partial_\mu + A_\mu \times) \tag{2.33}$$



which when operating on a vector produces a vector. We can now re-write (2.28) as

$$-\frac{1}{g^2} D_\mu E_{\mu\nu} = J_\nu \, . \tag{2.34}$$

Consider the action of the commutator $[D_\mu, D_\nu]$ on $\phi$. We easily see that

$$D_\mu D_\nu \phi - D_\nu D_\mu \phi = E_{\mu\nu} \times \phi \, . \tag{2.35}$$

This is the first time we have got this combination of $A$'s (viz. $E_{\mu\nu}$) out in a logical way.

Since $E_{\mu\nu}$ is an isovector, we can substitute it for $\phi$ in (2.35). Then we get, using the antisymmetry of $E_{\mu\nu}$

$$D_\mu D_\nu E_{\mu\nu} = 0 \, . \tag{2.36}$$

Comparing this with (2.34) we see that for consistency we must have

$$D_\mu J_\mu = 0 \, . \tag{2.37}$$

In other words these field equations are meaningless equations unless the current is conserved in the sense that its covariant divergence is zero. This causes a lot of complications when we go to the quantum theory. The reason is that in the quantum theory, when we calculate a diagram and so forth, some particles in the theory interact and provide a contribution to the current which is then a source which generates a new field propagating to the next interaction. We figure out how the vector fields propagate by solving the differential equations (2.34). It may not be that our source automatically satisfies (2.37), and hence eq. (2.37) does not make sense, it has no solution and we do not know how the field should propagate. Eq. (2.34) is the analogue of the Maxwell equations*

$$\partial_\mu F_{\mu\nu} = J_\nu \, . \tag{2.38}$$

The current $J_\mu$ is produced by the matter field and it is a consequence of the Dirac equation (2.27) (and its conjugate) that this current is indeed conserved and satisfies (2.37). If matter is a Dirac spinor as we have assumed, $J_\mu$ does not explicitly depend on $A_\mu$. This is not true in general, if we define $J_\mu$ as the

* The analogue of the other 2 Maxwell equations, which are an algebraic consequence of $F_{\mu\nu}$ being a curl, is

$$D_\mu E_{\rho\sigma} + D_\rho E_{\sigma\mu} + D_\sigma E_{\mu\rho} = 0 \, .$$

It is easy to check that this is an identity satisfied by $E_{\mu\nu}$ since it is of the form $\partial_\mu A_\nu - \partial_\nu A_\mu + A_\mu \times A_\nu$.

first variation of the matter term (in the action) with respect to $A_\mu$. The current would have a different form in, for example, scalar electrodynamics where the quadratic term is the Lagrangian has the form

$$(\partial_\mu + A_\mu)\phi^+(\partial_\mu + A_\mu)\phi + m^2\phi^+\phi . \tag{2.39}$$

The current is found by differentating this with respect to $A$ and is

$$J_\mu = \phi^+ \overleftrightarrow{\partial}_\mu \phi + 2\phi^+ A_\mu \phi , \tag{2.40}$$

and so in this case there is an extra non-matter term in the current. Let us write

$$E_{\mu\nu} = F_{\mu\nu} + A_\mu \times A_\nu , \tag{2.41}$$

where $F_{\mu\nu}$ looks like the field tensor in electrodynamics being just the curl of $A_\mu$. Then eq. (2.34) can be written as

$$-\frac{1}{g^2}\partial_\mu F_{\mu\nu} = J_\nu + \frac{1}{g^2}\{\partial_\mu(A_\mu \times A_\nu) + A_\mu \times E_{\mu\nu}\} . \tag{2.42}$$

This is just like electrodynamics. Each field is produced by a source; the source is isospin density (analogous to electric charge in electrodynamics) which here is a sum of the contributions from the matter and the field itself. The disadvantage of looking at it in this way is that we have lost the gauge invariance. It's strange but it's true that the amount of isospin density in the field depends upon the gauge – it's not a gauge invariant quantity. This is analogous to the way some people like to do gravitation.

The gravitational field equations (cf. (2.34)) are

$$G_{\mu\nu} = T_{\mu\nu} , \tag{2.43}$$

where $T_{\mu\nu}$ is the energy-momentum tensor of the matter and $G_{\mu\nu}$ is the Einstein tensor. Algebraically

$$D_\mu G_{\mu\nu} \equiv 0 , \tag{2.44}$$

where $D_\mu$ is some covariant derivative, and hence we have

$$D_\mu T_{\mu\nu} = 0 , \tag{2.45}$$

analogous to (2.37). However, as in (2.42) we can rewrite (2.43) as

$$G'_{\mu\nu} = T_{\mu\nu} + K_{\mu\nu} , \tag{2.46}$$

where $G'_{\mu\nu}$ is linear in $g_{\mu\nu}$ and $K_{\mu\nu}$ contains only terms quadratic and higher in $g_{\mu\nu}$. This is now the equation for a spin 2 particle where $K_{\mu\nu}$ is the energy-momentum density in the gravitational field, and we say that the gravitational field is produced by all energy, the energy of matter and the energy of the field itself; that's why it's non-linear. However if we make a generalised co-ordinate transformation, $T_{\mu\nu} + K_{\mu\nu}$ is not a real tensor; this is a famous problem, there is no real way to define the total energy-momentum tensor of the universe.



## 3. A geometrical look at gauge invariance

At each point in space-time imagine a frame defining axes in SU($n$) in some sense continuous (nearby frames nearly the same), but otherwise arbitrary. Then physically we might hope to define what we mean by the frames at $x$ and $x'$ are in the same direction. That is, imagine that we take an up particle (e.g. a proton in SU(2)) at $x$ and send it over to $x'$ so the guy at $x'$ can see what we call up. If there were no external influence, which could rotate isospin, acting in the space between $x$ and $x'$, we might hope to define and check that everyone is using the same frame and can expect that one choice of frames at all $x$ to be best in the sense of making the physics equations simplest (e.g. all parallel).

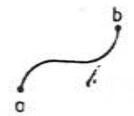
But under an influence, we must correct for the influence. If the influence is not universal (e.g. acts on protons but not on pions), we can compare frames using a particle which is affected least, or by looking at the different ways in which different particles are affected. If a universal influence acts which rotates the axis of isospin of every particle to the same degree, it clearly has no effect locally; but if the rotation varies from point to point and from time to time, than under some circumstances there may be an effect.

Now suppose there is such a universal influence and let us try to compare a frame at $b$ to a frame at $a$ by sending a particle from $a$ to $b$. Then the frame at $a$ "carried" to point $b$ might find (the frame as it is carried is of course turned by the universal influence), in general, that it is not lined up with the frame originally chosen at $b$, but requires an additional rotation $R(b \leftarrow a)$. Thus $R(b \leftarrow a)$ tells us how much the frame $b$ differs from the frame $a$ when $a$ is carried over, through space-time, to make the comparison. We could, of course, get a set of "best" frames so that all $R = 1$ by choosing at $b$ the frame we get by carrying our a frame to each space-time point.

$$R(c \leftarrow a) = R(c \leftarrow b)\, R(b \leftarrow a) , \tag{3.1}$$

i.e. unless $R(c \leftarrow a)$ is independent of the route by which a is carried, in which case we do not have an interesting physical theory at all; $R$ can be made equal to the identity by a proper choice of a universal set of frames.

The interesting theory arises if this is not the case, i.e. if the rotation $R(^{b \leftarrow a}_{\ell})$ depends on the path $\ell$ in space connecting the points. We now study this general case – the geometry of a field of frames with a method of parallel displacement, or comparison along geomatrical paths. (If the frames were Lorentz frames arbitrarily displaced, the theory is differential geometry and the physical theory is Einstein's General Relativity.)




Of course $R(^{b\leftarrow a}_{\ell})$ which compares frames at $b$ and $a$ does depend on the original frame choice at each point. If the frames at each point were rotated by $P(a)$ then the rotation between $b$ and $a$ would now be

$$R'\left(^{b\leftarrow a}_{\ell}\right) = P(b)R\left(^{b\leftarrow a}_{\ell}\right)P^{-1}(a)\,. \tag{3.2}$$

Physics should not depend on this choice, so we look for invariant properties of $R$, this is most easily done geometrically.

For a closed path $\ell_0$, $a \leftarrow a$, we might get a resulting rotation $R(^{a\leftarrow a}_{\ell_0})$ dependent on $\ell_0$. This determines a "strain" or physical effect independent of the choice of frames and thus invariant. To analyse these things most easily, we work with infinitesimal displacements and closed circuits (as any finite closed circuit can be represented as an area integral of infinitesimal closed circuits, in the manner familiar in the usual demonstration of Stokes's theorem).

Thus consider $b$ to be separated from $a$ by an infinitesimal coordinate displacement $\Delta x_\mu$. Then $R$ is nearly 1, the difference being of order $\Delta x_\mu$. Hence we can write, to first order in $\Delta x_\mu$

$$R(a + \Delta x \leftarrow a) = 1 - i\eta_\mu(x)\Delta x_\mu\,, \tag{3.3}$$

where $\eta_\mu$ is a vector field, in Minkowski space, depending on $x$, the location of point $a$, and is an operator in isospace. The transformation property of $\eta_\mu$ under a rotation $P$ of the frames is given (using (3.2)) by

$$1 - i\eta'_\mu \Delta x_\mu = P(x + \Delta x)(1 - i\eta_\mu \Delta x_\mu)P^{-1}(x)\,.$$

Putting

$$P(x + \Delta x) = P(x) + \frac{\partial P(x)}{\partial x_\mu}\Delta x_\mu$$

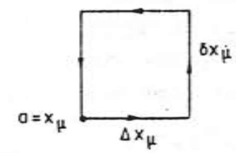
we get

$$\eta'_\mu = P(x)\eta_\mu P^{-1}(x) + i\frac{\partial P(x)}{\partial x_\mu}P^{-1}(x) \tag{3.4}$$



(we call this a gauge transformation of $\eta$).

What happens if we go round a small square?




(3.5)

Calculating to second order we obtain (to be correct to 2nd order we should expand each $R$ to second order beyond (3.3), but it is readily seen that these terms will cancel in this order in going up and down the sides of the square, i.e. such terms in the first bracket will cancel with terms in the third bracket, since to second order they are opposite)

$$R = \left[1 + i\eta_\mu\left(x + \frac{\delta x}{2}\right)\delta x_\mu + ...\right]\left[1 + i\eta_\nu\left(x + \frac{\Delta x}{2} + \delta x\right)\Delta x_\nu + ...\right]$$
$$\times\left[1 - i\eta_\sigma\left(x + \Delta x + \frac{\delta x}{2}\right)\delta x_\sigma + ...\right]\left[1 - i\eta_\tau\left(x + \frac{\Delta x}{2}\right)\Delta x_\tau + ...\right]$$
$$= 1 + i\{\partial_\mu\eta_\nu - \partial_\nu\eta_\mu + i[\eta_\mu, \eta_\nu]\}\,\delta x_\mu \Delta x_\nu \tag{3.6}$$
$$= 1 + i\mathcal{M}_{\mu\nu}\delta x_\mu \Delta x_\nu\,, \tag{3.7}$$

where we have defined

$$\mathcal{M}_{\mu\nu} = \partial_\mu\eta_\nu - \partial_\nu\eta_\mu + i[\eta_\mu, \eta_\nu]\,, \tag{3.8}$$

which is associated with an area $\delta x_\mu \Delta x_\nu$ and is an antisymmetric tensor operator of the second rank. $\mathcal{M}_{\mu\nu}$ is the physically interesting thing associated with the connection $\eta_\mu$ which takes us from place to place.

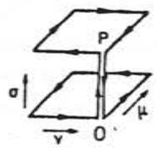
Suppose $\mathcal{B}$ is any tensor operator. We wish to know how it changes from place to place. It will not do simply to take $\mathcal{B}(x + \Delta x) - \mathcal{B}(x)$ since we cannot compare objects at a distance because of the effects of the universal influence. We must take account of the rotation of the frame by transporting $\mathcal{B}(x + \Delta x)$ back to $x$ before making a comparison. Hence the total change in $\mathcal{B}$ is

$$[1 + i\eta_\mu\Delta x_\mu]\mathcal{B}(x + \Delta x)[1 - i\eta_\nu\Delta x_\nu] - \mathcal{B}(x)$$
$$= \left[1 + i\eta_\mu\Delta x_\mu\right]\left(\mathcal{B}(x) + \frac{\partial \mathcal{B}}{\partial x_\tau}\Delta x_\tau\right)\left[1 - i\eta_\nu\Delta x_\nu\right] - \mathcal{B}(x)$$
$$= \left\{\frac{\partial \mathcal{B}(x)}{\partial x_\mu} + i[\eta_\mu, \mathcal{B}(x)]\right\}\Delta x_\mu\,. \tag{3.9}$$

This enables us to define a covariant derivative on any tensor $\mathcal{B}$ by

$$D_\mu \mathcal{B} = \partial_\mu \mathcal{B} + \mathrm{i}[\eta_\mu, \mathcal{B}] \,. \tag{3.10}$$

We will now deduce an interesting geometrical identity satisfied by $\mathcal{M}_{\mu\nu}$.




We wish to calculate the difference in circulation between the top and bottom faces of the cube. To get this difference it will not do simply to take $\mathcal{M}_{\mu\nu}(x + \mathrm{d}x_\sigma) - \mathcal{M}_{\mu\nu}(x)$, because to get back to $O$ we must go from $O$ to $P$ (a factor $1 - \mathrm{i}\eta_\tau \mathrm{d}x_\tau$), then around the top (a factor $1 + \mathrm{i}\mathcal{M}_{\mu\nu}(x + \mathrm{d}x)\delta x_\mu \Delta x_\nu$), and then back down to $O$ (a factor $1 + \mathrm{i}\eta_\sigma \mathrm{d}x_\sigma$); hence we want to compare

$$[1 + \mathrm{i}\eta_\sigma \mathrm{d}x_\sigma]\,[1 + \mathrm{i}\mathcal{M}_{\mu\nu}(x + \mathrm{d}x_\sigma)\delta x_\mu \Delta x_\nu]\,[1 - \mathrm{i}\eta_\tau \mathrm{d}x_\tau]$$

with

$$1 + \mathrm{i}\,\mathcal{M}_{\mu\nu}(x)\delta x_\mu \Delta x_\nu \,.$$

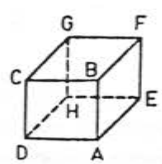
The difference between these two terms is

$$\left\{\frac{\partial \mathcal{M}_{\mu\nu}}{\delta x_\sigma} + \mathrm{i}[\eta_\sigma \mathcal{M}_{\mu\nu}]\right\} \mathrm{d}V_{\mu\nu\sigma} \,, \tag{3.11}$$

where

$$\mathrm{d}V_{\mu\nu\sigma} = \delta x_\mu \Delta x_\nu \mathrm{d}x_\sigma \,.$$

Notice that the term in the bracket is just the covariant derivative of $\mathcal{M}_{\mu\nu}$; this is not surprising as all we have done is to compare $\mathcal{M}_{\mu\nu}$ (cf. (3.5)) at $O$ and $P$.




By comparing $\mathcal{M}_{\mu\nu}$ on opposite pairs of faces, and noting that the net effect of going round the sum of the following 3 paths is zero,



A → E → H → D → C → G → F → B → C → D

~~(A → B → C → G → F → B → A → E → H → D → A)~~

+ (A → D → C → B → A → E → F → G → H → E → A)

+ (A → B → F → E → A → D → H → G → C → D → A) ,

we get

$$D_\sigma \mathcal{M}_{\mu\nu} + D_\mu \mathcal{M}_{\nu\sigma} + D_\nu \mathcal{M}_{\sigma\mu} = 0 \,. \tag{3.12}$$

By taking $B$ round a small closed loop, or by using (3.10), we get

$$D_\mu D_\nu B - D_\nu D_\mu B = \mathrm{i}[\mathcal{M}_{\mu\nu}, B] \,, \tag{3.13}$$

and hence

$$D_\mu D_\nu \mathcal{M}_{\mu\nu} = 0 \,. \tag{3.14}$$

Now we will associate this with what we did before with Yang–Mills theory. Here each $R$ is a rotation in isospace and so can be written in the form

$$R = \exp(-\mathrm{i}\boldsymbol{\alpha}\cdot \boldsymbol{T}) \tag{3.15}$$

where the $\boldsymbol{T}$ are the generators of some representation of the SU(2) Lie algebra, and $\boldsymbol{\alpha}$ is a vector depending on the rotation it is desired to represent.

For the infinitesimal rotation (3.3), $\boldsymbol{\alpha}$ is also an infinitesimal of first order in $\Delta x_\mu$, say $\boldsymbol{\alpha} = -\boldsymbol{A}_\mu \Delta x_\mu$ so

$$R = 1 + \mathrm{i}(\boldsymbol{A}_\mu \cdot \boldsymbol{T})\Delta x_\mu \,. \tag{3.16}$$

Comparing this with (3.3) we see that

$$\eta_\mu = -\boldsymbol{A}_\mu \cdot \boldsymbol{T} \,. \tag{3.17}$$

Substituting this into (3.8) we obtain

$$\mathcal{M}_{\mu\nu} = -\boldsymbol{E}_{\mu\nu} \cdot \boldsymbol{T} \tag{3.18}$$

where

$$\boldsymbol{E}_{\mu\nu} = \partial_\mu \boldsymbol{A}_\nu - \partial_\nu \boldsymbol{A}_\mu + \boldsymbol{A}_\mu \times \boldsymbol{A}_\nu \,, \tag{3.19}$$

and we see that is the same as the $\boldsymbol{E}_{\mu\nu}$ which we obtained in (2.24). The question now is – can we understand any of the Yang–Mills field equations (2.26)–(2.37) geometrically? The answer is for most of the equations, no. In particular, why did we put the combination $-(1/4g^2)\boldsymbol{E}_{\mu\nu}\cdot \boldsymbol{E}_{\mu\nu}$ in the Lagrangian? Physics tells us, not geometry. All we have discussed are the qualitative features of a classical field, and the response of particles to it (they rotate the axes as they move through it), but how the field itself has energy and behaves dynamically we haven't determined. But there are a few things we can work out. Using (3.12) and (3.18) we have

$$D_\mu E_{\nu\sigma} + D_\nu E_{\sigma\mu} + D_\sigma E_{\nu\mu} = 0 \,. \tag{3.20}$$

And using (3.13) and (3.14), we have

$$D_\mu D_\nu \bar{\phi} - D_\nu D_\mu \bar{\phi} = E_{\mu\nu} \times \bar{\phi} \,, \tag{3.21}$$

where $\bar{\phi}$ is any isovector and

$$D_\mu D_\nu E_{\mu\nu} = 0 \tag{3.22}$$

(cf. (2.35) and (2.36)).

The analogue of (2.34) would be

$$D_\mu \mathcal{M}_{\mu\nu} = g_\nu \,. \tag{3.23}$$

What is $D_\mu \mathcal{M}_{\mu\nu}$ geometrically? If I knew an easy way to describe this geometrically then we could state this equation as: $D_\mu \mathcal{M}_{\mu\nu}$ is the total isospin in a small volume. Unfortunately I haven't worked this out, and therefore I cannot describe the full Yang–Mills classical theory in an elementary way. (I am looking for a law like that in gravity, which says that the excess of the proper radius of a small 3 dimensional sphere over the radius calculated from the area, $\sqrt{(\text{area}/4\pi)}$, is proportional to the mass inside the sphere – which is, assuming Lorentz invariance, a complete statement of the Einstein law $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = T_{\mu\nu}$.)

We have said that the $A$ field represents a universal turning of the axes of a diffusing particle, and now we should check that the equation of motion of the matter $(i\not{\partial} - \not{A}\cdot\tau)\psi = m\psi$, for example, implies that indeed it does. To make it is easy, we first do it with the Schroedinger equation and electrodynamics but you will see that the method of proof is readily extendible to other cases. The free particle Schroedinger equation is

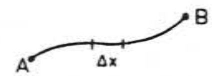
$$-\frac{\nabla^2\psi}{2m} = \mathrm{i}\frac{\partial\psi}{\partial t} \,. \tag{3.24}$$

Solving this equation, we find that a particle propagates as follows. Suppose it is confined at $x_1$ at time $t_1$ (the wave-function is a delta-function). Then the wave-function at time $t_2$ and position $x_2$ is

$$\left[\frac{2\pi\mathrm{i}(t_2 - t_1)}{m}\right]^{-3/2} \exp\left[\frac{\mathrm{i}m(x_2 - x_1)^2}{2(t_2 - t_1)}\right] , \tag{3.25}$$

which we call $K_0(2,1)$, this being the function which describes how the particle diffuses outwards in time. Now consider a possible trajectory for the particle






The amplitude for the particle to go from $A$ to $B$ can be found by multiplying the amplitudes for it to go along all the infinitesimal sections of the path.

Now add an external field to the Schroedinger equation

$$\frac{1}{2m}(\mathrm{i}\underline{\nabla} - \underline{A})^2\psi + V\psi = \mathrm{i}\frac{\partial\psi}{\partial t} \tag{3.26}$$

Consider an infinitesimal distance $\Delta x$. Over this distance, $A$ and $V$ may be treated as constants and are therefore expressible as the gradient of a potential $\chi$

$$\underline{A} = \underline{\nabla}\chi, \qquad V = \partial\chi/\partial t \,, \tag{3.27}$$

where

$$\chi = \underline{A}\cdot\underline{x} + Vt \,.$$

By a gauge transformation we see that the wave-function $\psi' = \mathrm{e}^{-\mathrm{i}\chi}\psi$ is a solution of (3.26) if $\psi$ is a solution of (3.24), and hence we may write the amplitude for propagation over a short distance and time $\Delta x$ in the presence of the field as

$$K_A(x + \Delta x, x) = \exp\{\mathrm{i}\chi(x + \Delta x)\}\exp\{-\mathrm{i}\chi(x)\}\, K_0(x + \Delta x, x)$$

$$= \exp(\mathrm{i}A_\mu \Delta x_\mu) K_0(x + \Delta x, x) \,, \tag{3.28}$$

where

$$A_0 = -V, \qquad A_i = \underline{A} \,.$$

Iterating this over a continuous path, we get

$$\exp(\mathrm{i}\int A_\mu \mathrm{d}x_\mu) \times (\text{Amplitude for process without the } A \text{ field}) \,. \tag{3.29}$$

In the case of Yang–Mills theory, this generalises to

$$\exp(\mathrm{i}\int A_\mu \cdot T \mathrm{d}x_\mu) \times (\text{Amplitude for process without Y–M field}) \,, \tag{3.30}$$

where we must order the operator $T$ along the path

For infintesimal paths this reduces to

$$[1 + i(A_\mu \cdot T)\Delta x_\mu] \times (\text{Amplitude for process without Y–M field}) , \quad (3.31)$$

and thus we see that the expression (3.16) has emerged naturally.

## 4. A qualitative critique of QCD

One of the possible applications of Yang–Mills theory is to Quantum Chromodynamics. The question to which we wish to address ourselves in this chapter is whether or not this theory has a real chance of being right. Ordinarily, when the right theory is found, it isn't long before we can calculate consequences and check that it agrees with everything relevant that is known. For example, the whole subject of electrostatics was in complete confusion until the Coulomb law was discovered; before then people were rushing around in complete chaos and then suddenly they were calculating the capacity of elliptical condensors etc. ... . Similarly, before Schroedinger's equation, there was a lot of pulling and hauling on ideas which were inconsistent, and suddenly, as soon as the equation was discovered, there was a tremendous tumbling out of results which showed how everything worked. Therefore it's expected (at least by an old fogey like myself) than when the correct theory is found, lots of results will tumble out which will agree with experiment. Now QCD is proposed as a theory which is supposed to be the correct theory of strong interactions; it's been around for a few years now and we don't have any quantitative results. At the moment we cannot look at the theory quantitatively (due possibly to technical difficulties in its interpretation) so we will look at it qualitatively to try and decide whether it is useful or not.

### *4.1. Forces between the quarks*

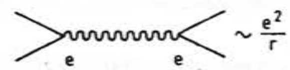
The most characteristic thing about the quark bound state which have been seen is that they are colour singlets. Coloured states have not been seen and we must conclude that either their mass is infinite or is out of the reach of present experiments.

In this theory the three quarks in a baryon form an antisymmetric state with respect to colour:

$$\frac{1}{\sqrt{6}}\{|A\rangle|B\rangle|C\rangle + |B\rangle|C\rangle|A\rangle + |C\rangle|A\rangle|B\rangle - |B\rangle|A\rangle|C\rangle - |A\rangle|C\rangle|B\rangle - |C\rangle|B\rangle|A\rangle\}. \quad (4.1)$$

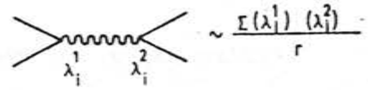


The first thing to look at is when we put three quarks together, can their energy be lower in some other state than in the singlet state? Similarly putting a quark and an anti-quark together, is the colourless (singlet) state lower than any other state? At the level we will work, we will not attempt to calculate the energies correctly, we merely wish to see whether we can get their order right. In QED we know that the force between two static particles is




(4.2)

Similarly the force between two static quarks due to single gluon is




(4.3)

where $\lambda_i$ are the SU(3) matrices.

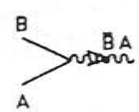
One can see this as follows. For a static source $J_x = J_y = J_z = 0$, and $J_t = \rho$, the density of colour charge, we find a solution of the form $A_x = A_y = A_z = 0$, $A_t \neq 0$ where

$$\nabla^2 A_t = \rho_t \quad (4.4)$$

(from 2.34). This is just as in electrostatics except that here we have three isospin components. We could look at this problem mathematically (by fiddling around with the $\lambda$'s); however we wish to take a simpler view. The gluons which couple to the colours of the quarks, can be represented as

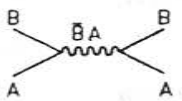
$$\bar{A}B,\ \bar{B}A,\ \bar{A}C,\ \bar{C}A,\ \bar{B}C,\ \bar{C}B,\ \frac{1}{\sqrt{2}}(\bar{B}B - \bar{C}C),\ \frac{1}{\sqrt{6}}(2\bar{A}A - \bar{B}B - \bar{C}C). \quad (4.5)$$

The last two gluons are the non colour changing states orthogonal to the singlet $(1/\sqrt{3})(\bar{A}A + \bar{B}B + \bar{C}C)$ which is omitted for reasons discussed earlier. This symbolism really tells us what happens to the colour of a quark when it emits or absorbs a gluon. That is, an $\bar{A}B$ gluon can be absorbed by an A quark turning it into B, with amplitude 1. (Annihilate the A and create a B instead.) We now calculate the relative strengths with which various quarks couple. Consider this vertex; an A quark emits a $\bar{B}A$ gluon changing to B.



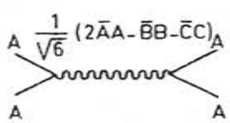

(4.6)

This $\bar{B}A$ gluon can be absorbed by a B quark (turning it to A) but cannot be absorbed by an A or C (it can by an $\bar{A}$ antiquark however). Therefore, if we wish to look at the A–A quark force, this vertex cannot contribute because this gluon cannot be absorbed by an A quark, i.e.




does not go.

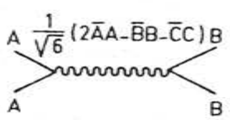
The only contribution to the A–A force is that due to the colourless gluon exchange, viz.




(4.7)

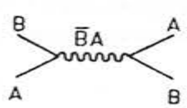
The interaction energy for this process is $+2/\sqrt{6}\,(+2/\sqrt{6}) = +\frac{2}{3}$, which is positive, so they repel just as in electrostatics (like charges repel).

Now what about the A–B force? Here there are two allowed diagrams.




$$\text{interaction energy} = +\frac{2}{\sqrt{6}}\left(-\frac{1}{\sqrt{6}}\right) = -\frac{1}{3}, \quad (4.8)$$

and the exchange diagram




$$\text{interaction energy} = (+1)(+1) = +1\,. \quad (4.9)$$

4.8 and 4.9 must be combined to give the total interaction energy which gives $+\frac{2}{3}$.

We could do the same for all the other colour combinations, but we'd be wasting our time as we know that the interaction is symmetric with respect to colour. The only cases we have to worry about are when the colours of the quarks are the same or different. We can summarize the results as follows. Introduce a colour exchange operator $P$ which interchanges a pair of quarks, it



has eigenvalues +1 or −1 corresponding to whether the quarks are in a symmetric or antisymmetric colour state. In the case when the quarks are different (cf. (4.8), (4.9), the interaction energy can be written

$$(p - \tfrac{1}{3})\,, \quad (4.10)$$

where $p$ is the eigenvalue of $P$. The lovely thing about this formula is that, if we apply it to the case where the colours are the same (cf. (4.7), $p = 1$ in this case), we get the right answer $1 - \frac{1}{3} = \frac{2}{3}$. Thus we have shown that $\Sigma(\lambda_i)^1(\lambda_i)^2 = P - \frac{1}{3}$. This is analogous to the Dirac formula for the interaction between two spins (e.g. two electrons in an atom):

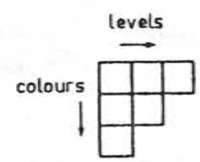
$$p_{\text{exch}} - \tfrac{1}{2} = \sum_i (\sigma_i)^1(\sigma_i)^2\,, \quad (4.11)$$

(the $-\frac{1}{2}$ becomes $-1/n$ for SU($n$).

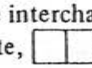
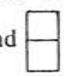
In order to generalise formula (4.10) to states of more than two quarks, we merely sum it over all possible pairs of quarks.

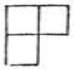
We shall now calculate the energy of various quark states using (4.10). Each quark will have some self-energy, we don't know what this is, but to make it easier to see what's going on, we will take $+\frac{4}{3}$ for each quark – the qualitative results do not depend on this choice (because we will always compare the energies of states of the same total number of quarks).

We will symbolise the quark states as follows: draw a series of boxes with one box for each quark in the state, for example a possible 6 quark state is




(4.12)

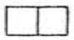
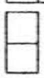
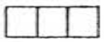
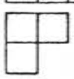
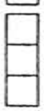
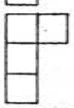
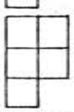
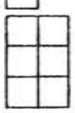
This represents a quark configuration in which the wave-function is symmetric with respect to the interchange of any pair of quarks in the same row, and antisymmetric with respect to the interchange of any pair of quarks in the same column, e.g. for a two quark state,  represents the symmetric state and  represents the antisymmetric state. We can use these (Young) diagrams to calculate the energy of a bound state using (4.10). We will work out one of these diagrams in detail for a three quark state which is symmetric under exchange of one pair of quarks and antisymmetric under exchange of another pair:

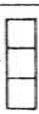
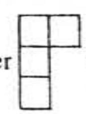
 (4.13)

The contribution from the exchange is

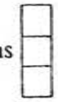
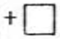
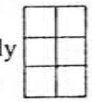
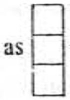
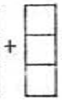
$$\underset{\text{row}}{(+1)} + \underset{\text{column}}{(-1)}\,.$$

Therefore the total interaction energy is $(+1) + (-1) - 3(\frac{1}{3}) = -1$. When we add $+\frac{4}{3}$ for each quark we obtain $+3$. The following table summarises the results for various states.

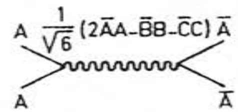
| No. of quarks | Diagram | Interaction energy | State energy ($+\frac{4}{3}$ added per quark) |
|---|---|---|---|
| 1 |  | $0$ | $+\frac{4}{3}$ |
| 2 |  | $+1 - \frac{1}{3} = +\frac{2}{3}$ | $+\frac{10}{3}$ |
| |  | $-1 - \frac{1}{3} = -\frac{4}{3}$ | $+\frac{4}{3}$ |
| 3 |  | $3(+1) - 3(\frac{1}{3}) = +2$ | $+6$ |
| |  | $+1 + (-1) - 3(\frac{1}{3}) = -1$ | $+3$ |
| |  | $3(-1) - 3(\frac{1}{3}) = -4$ | $0$ |
| 4 |  | $+1 + 3(-1) - 6(\frac{1}{3}) = -4$ | $+\frac{4}{3}$ |
| 5 |  | $2(+1) + 4(-1) - 10(\frac{1}{3}) = -\frac{16}{3}$ | $+\frac{4}{3}$ |
| 6 |  | $3(+1) + 6(-1) - 15(\frac{1}{3}) = -8$ | $0$ |

Notice that , which is totally antisymmetric with respect to colour and which we identify with a baryon, has the lowest energy. Further  has the



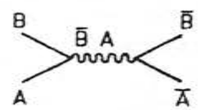
same energy as  +  i.e. a single quark should be only weakly bound to a proton. Similarly  has the same energy as  +  (in our approximation, zero) and it is impossible to say, in this poor approximation, whether or not such a bound state of two baryons will exist.

Now let us try the same thing for the mesons. Consider



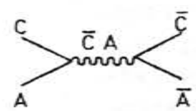

$$\text{interaction energy} = +\frac{2}{\sqrt{6}}\left(-\frac{2}{\sqrt{6}}\right) = -\frac{2}{3}. \qquad (4.14)$$

The minus sign appears at the antiparticle vertex because the theory is a vector theory; this is just like electricity where the antiparticle has the opposite charge to the particle. Another possible term in the A–Ā interaction is that in which the quarks change their states




$$\text{interaction energy} = +1\,(-1) = -1 \qquad (4.15)$$

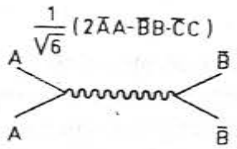
Similarly




has the same energy as (4.15), $-1$. (4.16)

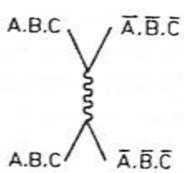
Now we can figure out the energy of the meson state $(1/\sqrt{3})(|A\rangle|\bar{A}\rangle + |B\rangle|\bar{B}\rangle + |C\rangle|\bar{C}\rangle)$ (singlet) i.e. we need to consider the coupling of this state to itself. Since this state is symmetric in A, B, C we get $3 \times 1/\sqrt{3}|A\rangle|\bar{A}\rangle \times 1/\sqrt{3}(|A\rangle|\bar{A}\rangle + |B\rangle|\bar{B}\rangle + |C\rangle|\bar{C}\rangle)$ which is simply the sum of (4.14)–(4.16). Therefore the interaction energy is $-\frac{8}{3}$. Adding $+\frac{4}{3}$ for each quark, we find the meson energy to be 0. (The mesons have the same energy scale as baryons, with this $+\frac{4}{3}$ choice for each quark.)

Finally, we examine the state $|A\rangle|\bar{B}\rangle$, a coloured meson: there is only one possible diagram




$$\text{interaction energy} = +\frac{2}{\sqrt{6}}\left(+\frac{1}{\sqrt{6}}\right) = +\frac{1}{3}. \qquad (4.17)$$

Adding the quark self-energy we get a coloured meson energy of +3, which is much higher than for a colourless meson. We could ask why




does not contribute to the energy of the colourless meson. The answer is that the $s$ channel gluon must be colourless; no such gluon exists – it was eliminated earlier.

### *4.2. Infra-red behaviour*

We have indicated that a $r^{-1}$ potential exists between two quarks in this theory. We know that this cannot really be correct, since the force between two quarks is known not to be long range – it is not easy to knock the quarks apart inside a proton (c.f. the case with which we can knock electrons out of atoms). This problem cannot be argued away by saying that the charges in QCD are stronger than in QED – with enough energy, we should be able to pull them apart. The long distance (infra-red) forces have to be modified in order to agree with experiment. Those who believe in QCD believe that this will happen when the theory is worked out. I think that the central problem in QCD is to see if, qualitatively, the forces are so modified. In some models, e.g. the lattice theory (see [12]), they are, but we are not sure if this an artifact of the models or not. It is true that when we go to higher order perturbations, we find the force increasing with distance (there are logarithmic corrections) relative to $1/r^2$, but it is unknown if this change is sufficient.



### *4.3. Dependence of masses on flavours*

#### *4.3.1. Isospin dependence*

The proton and the $\Delta$ are both supposed to be made out of 3 quarks in a totally antisymmetric colour state.

| | Mass (MeV) | Isospin | Spin |
|---|---|---|---|
| $P$ | 938 | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $\Delta$ | 1236 | $\frac{3}{2}$ | $\frac{3}{2}$ |

If QCD has forces depending only colour, how could there be a difference between these two masses depending on the isospin? In addition to the isospin being different, the spins are also different and we know, for example, that the forces are spin dependent and different for different spin states in QED; so there is no reason why this cannot be the case in QCD. The difference in the $P$ and $\Delta$ masses, therefore, may simply be due to the fact that their spins are different. Then we say – look through the Rosenfeld tables to find particles with different isospins and masses, but the same spins and for which we expect the same space. Then QCD could be in trouble. However we cannot find any for the following reason.

The wave-function is antisymmetric with respect to interchange of two quarks (Fermi statistics). This total exchange is equivalent to an exchange of space, spin, flavour and colour. Since the state is antisymmetric with respect to colour exchange (colour singlet), the flavour symmetry must equal the space symmetry × the spin symmetry; hence the flavour symmetry properties of a state are completely determined by its space and spin symmetry properties. But colour forces depend on the space and spin configurations and therefore can apparently depend on the isospin symmetry. In other words, we can find no way to verify the proposition that the forces are independent of flavour. This, by the way, should be noted because in the early days of hadron theory the forces had explicit isospin dependence, e.g. there were interaction terms like $\bar{\psi}(\phi\cdot\gamma)\psi$, for a proton coupled to a pion. In QCD we cannot have any such directly isospin dependent terms, but it doesn't matter as we have seen that the masses can depend indirectly on isospin.

A very rough empirical formula exists [13] which summarises the mass splitting between baryon multiplets. It says that there is a contribution of $-0.53\ (\text{GeV})^2$ to the $(\text{mass})^2$ for every pair of quarks which are both symmetric in space and antisymmetric in spin. Mathematically this is

$$\Delta M^2 = -0.53\ (\mathrm{GeV})^2 \sum_{\text{pairs}} \left(\frac{1 - P_{\text{ex spin}}}{2}\right)\left(\frac{1 + P_{\text{ex space}}}{2}\right). \tag{4.18}$$

Notice that the masses are lower if the spacial state is symmetric as opposed to antisymmetric, in an antisymmetric state, it is impossible for the quarks to be at the same point; but in the symmetric state this is allowed, so that the qualitative features of (4.18) may be summarized by saying that the correction force is short range. The spin part says that antiparallel spin states are lower which is the right sign for the spin interaction of attracting particles due to a vector potential.

*4.3.2. Dependence on the quark masses*

We know that the K mass does not equal the $\pi$ mass, although the K and the $\pi$ both have the same spin-parity. How can we explain this? If all the forces are independent of flavour, then the masses should be exactly the same. It is therefore a failure of the simplest possible picture that SU(3) is broken. But we do not need to destroy QCD, we need only complicate it by supposing that the s quark has a higher mass than the u and d quarks (but keeping the interaction independent of flavour). No-one knows where this extra mass comes from, we have left for the people of the future the problem of why the masses are different. Some people think that the masses of the u and d quarks are equal and that all the mass differences in isospin multiplets are due to electrodynamics (e.g. proton–neutron), there are some technical difficulties in signs and magnitudes in this approach and it would help if we could say that the d quark is slightly heavier than the u quark for the same intrinsic (unknown) reason that the s quark is heavier than the u and d quarks. Another flavour now seems to have been found and maybe there are more; this new quark (usually called charm) must have a higher mass than the other three. We mustn't forget that it is strange that we have to put in mass differences which are of the same order of magnitude as the bound state masses we are trying to explain. If we go to very high momentum transfers where masses are irrelevant, the SU(3) (flavour) should become better, I don't know of any direct demonstration that this is, or is not, the case – it would be nice to think of some experiment in which this could be tested.

*4.4. Zweig rule*

Consider the following groups of mesons



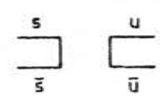
| | $\pi$ | $\eta$ | $\eta'$ | $0^-$ |
|---|---|---|---|---|
| Mass(GeV) | 0.14 | 0.55 | 0.96 | |
| | $\rho$ | $\omega$ | $\phi$ | $1^-$ |
| Mass(GeV) | 0.77 | 0.78 | 1.02 | |
| | $A_2$ | f | f′ | $2^+$ |
| Mass(GeV) | 1.31 | 1.27 | 1.52 | |

These are the non-strange mesons from the $0^-$, $1^-$ and $2^+$ nonets and one might expect some similarity between these groups. Now it is known that the $\phi$ (e.g. in its decay, $K\bar{K}$ dominates) is an $|s\bar{s}\rangle$ state. Similarly the $\omega$ ($3\pi$ decay mode dominates) is $1/\sqrt{2}(|u\bar{u}\rangle + |d\bar{d}\rangle)$ and the $\rho$ is $1/\sqrt{2}(|u\bar{u}\rangle - |d\bar{d}\rangle)$ and they are very nearly degenerate. But the $\eta$ and the $\eta'$ do not have this pattern Approximately,

$$\eta = \tfrac{1}{2}(|u\bar{u}\rangle + |d\bar{d}\rangle - \sqrt{2}|s\bar{s}\rangle) \quad \text{and} \quad \eta' = \tfrac{1}{2}(|u\bar{u}\rangle + |d\bar{d}\rangle + \sqrt{2}|s\bar{s}\rangle)$$

agrees with experiment. Note that this combination is very different from the $\omega$, $\phi$ case. Why? We would think that, in a state which contains $|s\bar{s}\rangle$, the s and $\bar{s}$ could annihilate and then turn back into s and $\bar{s}$, but also into u and $\bar{u}$ or d and $\bar{d}$; therefore an $|s\bar{s}\rangle$ state should become a mixture of all 3 quark types, and this is presumbly what is happening in the $\eta$, $\eta'$ system. Why does this not happen in the $\omega$, $\phi$ case (and also does not happen in the f, f′ case)? This is a mystery. This mystery is summed up in the Zweig rule, an ad hoc rule which says that this annihilation process is inhibited.

One possible attitude is as follows




With an $0^-$ state, we need 2 gluons to connect both sides of this diagram. At first sight, we would think that with a $1^-$ state, we could connect with 1 gluon since the gluon has quantum numbers $1^-$ (cf. photon); but this would require a colour singlet gluon and we have no such object; so we need a minimum of 3 gluons. If we could assume that the gluons were weakly coupled to the quarks, we might try to argue that the mixing in the $\omega$, $\phi$ case is less than in the $\eta$, $\eta'$ case, but it's a little hard to get $g^3$ less than $g^2$ unless $g$ (the coupling constant) itself is very small. Furthermore, if we look at the $2^+$ mesons, we could make the connection in the diagram with only 2 gluons, so it should look rather like $\eta$, $\eta'$ system; but it doesn't (here the suggestion was made that the higher angular momentum $2^+$ makes it harder for the gluons to get together to annihilate).

There is also another attitude, which is to suppose that, for some reason, when momentum transfers and energies become large, the coupling becomes small, then we can say the mixing is less in the $2^+$ case than in the $0^-$ case because the energy is higher. In order for this work, the rate of change of coupling with mass scale must be large.

This is not a satisfactory situation – it looks suspicious, as there is some feature of QCD that we do not understand. If we have a way of calculating with QCD (e.g. on a lattice) and if we wish to concentrate on something which will tell us if the theory is wrong, this seems to me to be an ideal place.

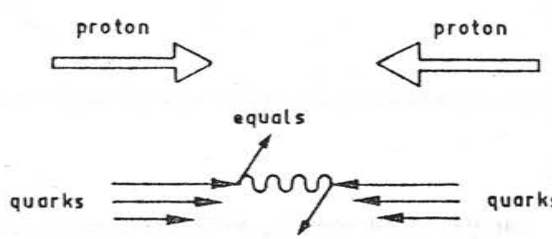
### *4.5. Large transverse momentum in hadronic collisions*

Consider the process

$$p + p \to \text{hadron (large } p_\perp) + X \, . \tag{4.19}$$

The average $p_\perp$ of a produced hadron at high energies is of order 350 MeV. In a field theory, it is not easy to understand why there isn't a reasonable amount of larger $p_\perp$. Let us look specifically at particles produced at very large $p_\perp$.

In our picture the proton is made of a bunch of quarks, so we might expect that a large $p_\perp$ particle would be produced by quarks scattering via gluon exchange.




(4.20)

If the ratio $x_\perp = p_\perp/p$ is kept constant, the cross-section at large $p_\perp$ should go as $p_\perp^{-4}$. (This follows merely by dimensional arguments and this result would be obtained for any gluon diagram.) Experimentally, it is more like $p_\perp^{-8.2}$. This is very disturbing. Is this process operating or not? A possibility is that it really does occur, but when we put in the correct couplings and allow for



the fact that the coupling constant will fall as $p_\perp$ increases, the process (4.20) is masked by some other process which falls like $p_\perp^{-8.2}$. We must then assume that this other process hasn't yet fallen enough to let us see the mechanism (4.20). However the cross-section is already pretty small at 400 GeV and still seems to be falling like $p_\perp^{-8}$; it is therefore up to the people who propose this explanation to give some energy above which they believe the process (4.20) will take over; and to explain what mechanism is responsible for the present trend. (One possible mechanism is described by the constituent interchange model, which gives a cross-section falling like $p_\perp^{-8}$.) But in any event there is a challenge to see what process QCD predicts that is so large as to dominate over all the experimental range so far investigated and which behaves like $p_\perp^{-8}$. Any theory of this process has to explain much data such as charge ratios correlations, etc… . Details such as the fact that as $x_\perp$ rises (toward 0.6) the ratio $\pi^+/\pi^-$ rises to more than 2, have to be explained.

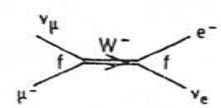
### *4.6. Partially conserved axial current and vector meson dominance*

Certain hadrons have special properties and it is not clear where these properties come from. One of these hadrons is the pion, which has the property of PCAC associated with it; another is the $\rho$ meson which participates in VMD.

Both PCAC and VMD were discovered when it was thought that the particles were themselves fundamental fields. Why bound states of quarks (as QCD assumes the $\pi$ and $\rho$ are) should behave as if they were fundamental fields is a puzzle. I don't know whether this is a serious problem for QCD or not.

## 5. Spontaneously broken symmetries and the Higgs mechanism

We will try to show why it is necessary to use a Yang–Mills theory in weak interactions by considering the difficulties which appear in a more phenomenological approach. We will discuss some relevent points from weak interaction theory although we do not propose to give a review of it. Consider the decay of a $\mu^-$; one interpretation of this decay is the following diagram.




(5.1)

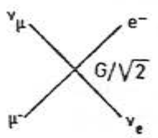
where the weak interaction is mediated by a massive charged spin one boson $W^-$. (The $W^-$ must be massive as the weak interaction is short range.) The couplings at the two vertices are assumed to be V–A and equal (coupling constant $f$). The amplitude for this decay is

$$\alpha = \frac{f^2}{q^2 - M_W^2}\,, \tag{5.2}$$

where $q$ is the momentum transfer between the $\mu^-$ and the $e^-$. For $q \ll M_W$, this reduces to

$$f^2/M_W^2\,. \tag{5.3}$$

In this limit the interaction looks like the four point Fermi interaction




$$\text{Amplitude} = (G/\sqrt{2}) \tag{5.4}$$

i.e. comparing (5.3) with (5.4) we get

$$f^2/M_W^2 = G/\sqrt{2}\,. \tag{5.5}$$

Let us write a Lagrangian for the $W^-$ and its anti-particle, the $W^+$, with a mass; we do this by analogy with QED for a massive photon:

$$\mathcal{L} = -\tfrac{1}{4}(\partial_\mu W_\nu - \partial_\nu W_\mu)^2 + \tfrac{1}{2}M^2 W_\mu W_\mu + \text{matter terms}\,. \tag{5.6}$$

Varying the action with respect to $W_\mu$ to obtain the equations of motion, we get

$$\partial_\mu(\partial_\mu W_\nu - \partial_\nu W_\mu) + M^2 W_\nu = S_\nu\,, \tag{5.7}$$

where $S_\mu$ is the current generated by the matter fields in the Lagrangian. $S_\mu$ is not conserved: one contribution to $S_\mu$ (the $\mu\nu W$ vertex in (5.1)) is

$$\bar{\psi}(\nu_\mu)\gamma_\mu(1-\gamma_5)\psi(\mu)\,. \tag{5.8}$$

This has non-zero divergence. It would have zero divergence if the $\gamma_5$ were absent and the masses of $\mu^-$ and $\nu_\mu$ were equal, but this non-zero divergence is all right since the divergence of (5.7) is

$$M^2 \partial_\mu W_\mu = \partial_\mu S_\mu\,. \tag{5.9}$$

So we see that the presence of the mass term saves us from a possible inconsistency.

We derive the propagator for the $W_\mu$ as follows. Eq. (5.7) can be re-written using (5.9) as



$$\Box W_\mu + M^2 W_\mu = S_\mu + \frac{1}{M^2}\partial_\mu(\partial_\nu S_\nu)\,. \tag{5.10}$$

In momentum space, this is

$$W_\mu = \frac{1}{k^2 - M^2}\left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{M^2}\right)S_\nu\,. \tag{5.11}$$

Therefore, given the sources $S_\nu$, we can calculate the field $W_\mu$ using the propagator

$$\frac{\delta_{\mu\nu} - (k_\mu k_\nu/M^2)}{k^2 - M^2}\,. \tag{5.12}$$

This tells us how a field $W_\mu$ propagates from one interaction to the next. Using this, we can construct diagrams for the theory. There are indices on the propagator because $W$ is a vector and therefore has various polarisation states. We can see that a free $W_\mu$ has three polarisation states, as required of a vector particle, and not four as we might guess at first sight: take (5.7) with $S_\mu = 0$; this describes the propagation of non interacting W's. Substitute a free particle solution

$$W_\mu = e_\mu\, e^{ik\cdot x}\,, \tag{5.13}$$

where $e_\mu$ is the $W$ polarisation vector. This gives

$$-k_\mu(k_\mu e_\nu - k_\nu e_\mu) + M^2 e_\nu = 0\,. \tag{5.14}$$

Multiplying by $k_\nu$ we get

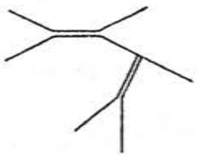
$$-k^2(k\cdot e) + k^2(k\cdot e) + M^2(k\cdot e) = 0\,. \tag{5.15}$$

Hence

$$k\cdot e = 0 \quad (M \neq 0)\,, \tag{5.16}$$

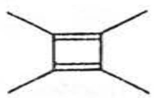
i.e. the polarisation vector $e_\mu$ is orthogonal to $k_\mu$ and so has only three degrees of freedom. Of course, when the W's are off mass shell, there are four polarisation states. Similarly, in massless QED, the photon has two polarisation states when it is on its mass shell, and three when it's off (the extra degree of freedom represents the freedom to make gauge transformations in QED which is lost if a mass term is added as in (5.6)).

We attempt to calculate with this massive W theory. Everything works well as long as we do not have any diagrams with closed loops i.e. as long as we only calculate tree diagrams:



(5.17)

Another way to say this is that the theory works perfectly at the classical level. At the quantum level, we run into difficulties with closed loops



(5.18)

The reason is that we must integrate over the momentum $k$ running round the closed loop. At large $k$, the propagator, 5.12 $\sim k_\mu k_\nu / k^2$ so the propagator does not assist the convergence of the integral; the diagram diverges. Unfortunately we cannot remove these divergences as in QED, because as we go to more loops the divergences become more and more severe: (i.e. it is not renormalizable), e.g.


(5.19)

The theory is therefore a disaster quantum mechanically and in order to construct a workable renormalisable theory of weak interactions we go to a Yang–Mills theory with broken symmetry. To do this, we first discuss symmetry breaking in simpler cases. I shall take a rather physical view of symmetry breaking and leave the more abstract mathematics to other people. I do this because I think that it is useful to have more than one way of looking at a problem; the way I shall present it may be unfamiliar, but people may benefit from this physical approach.

We will start with a simple model and gradually increase its complexity. First take a real scalar field only; the Lagrangian is

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu \phi)^2 - \frac{\mu^2}{2}\phi^2 - \frac{\lambda}{4}\phi^4 , \tag{5.20}$$

where we have included a $\phi^4$ interaction term. The theory has discrete symmetry $\phi \to -\phi$. Normally the vacuum expectation value of $\phi(\langle\phi\rangle)$ is zero; however



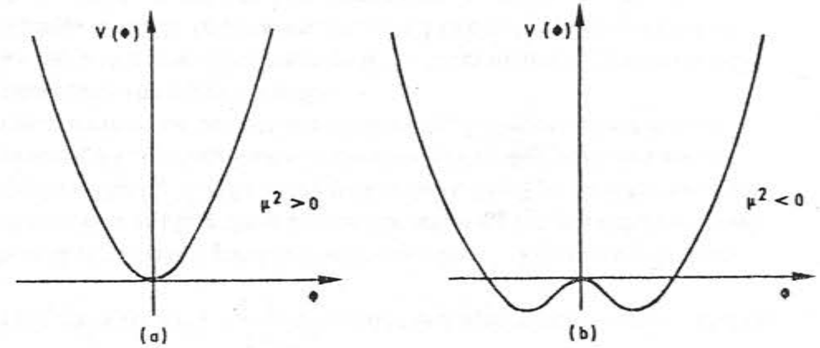



Fig. 5.1.

this is not always the case as we must choose $\langle\phi\rangle$ to minimise the energy.

In fig. 5.1, we have plotted the classical potential

$$V(\phi) = \frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \tag{5.21}$$

for the two cases (a) $\mu^2 > 0$ and (b) $\mu^2 < 0$. We look for the state of minimum energy (the vacuum) which we get by minimising $V(\phi)$:

$$\partial V(\phi)/\partial\phi = 0 , \tag{5.22}$$

i.e. $\phi = 0$ or $\phi = \pm\sqrt{-\mu^2/\lambda}$.

From fig. 5.1, we see that

$$\text{for} \quad \mu^2 > 0, \quad \langle\phi\rangle = 0$$

$$\text{for} \quad \mu^2 < 0, \quad \langle\phi\rangle = \pm\sqrt{-\mu^2/\lambda} = \pm v . \tag{5.23}$$

Take $\langle\phi\rangle$ to be $+v$, although the choice of sign is arbitrary; however notice that once we have chosen $\langle\phi\rangle = +v$, we have broken the symmetry $\phi \to -\phi$.

In case (b), we now perturb about $v$

$$\phi(x) = v + \eta(x) \tag{5.24}$$

Substituting this in (5.20) we get

$$\mathcal{L} = -\tfrac{1}{2}(\partial_\mu \eta)^2 - \lambda v^2 \eta^2 - \lambda v \eta^3 - \frac{\lambda}{4}\eta^4 + \frac{\lambda}{4}v^4 . \tag{5.25}$$

Suppose $\eta$ is small*; to a first approximation only the first two terms matter; these represent a scalar meson with a real mass of $\sqrt{2\lambda\, v^2}$. (Note that in 5.20 the $\phi$ meson had an imaginary mass in the case $\mu^2 < 0$.) The other two $\eta$ terms in the Lagrangian represent $\eta$ self couplings viz.


(5.26)

These couplings would appear when we do perturbation theory with (5.25).

Why does the whole world have $\langle\phi\rangle = +v$? Why doesn't it have $\langle\phi\rangle = -v$ somewhere? Suppose that God created the universe in the state $\langle\phi\rangle = 0$ and then the universe discovered that it could lower its energy; where it puts its energy is none of my business, but it gets rid of it – gives it back to God or something; then under some disturbance the vacuum tries to fall down with some parts going to $+v$ and other parts to $-v$. But what happens in between? It just changes suddenly, but not too suddenly, because to get low energy the $(\partial_\mu\phi)^2$ term must not be too large. This extra $(\partial_\mu\phi)^2$ energy, plus the energy due to the fact that $\phi$ is above its minimum potential energy, is stored in the boundary between the two regions, so that if we have some region in which $\langle\phi\rangle = -v$ surrounded by a region where $\langle\phi\rangle = +v$, we can lower the energy by shrinking the region where $\langle\phi\rangle = -v$ to zero decreasing the surface area and hence the energy so that the whole universe is in the same state.

Secondly, we look at a complex scalar field i.e. $\phi$ is a two component field with the components Re $\phi$ and Im $\phi$.

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\phi)^*(\partial_\mu\phi) - \frac{\mu^2}{2}\phi^*\phi - \frac{\lambda}{4}(\phi^*\phi)^2 , \qquad (5.27)$$

with $\mu^2 < 0$.

We minimise the potential

$$V(\phi,\phi^*) = \frac{\mu^2}{2}\phi^*\phi + \frac{\lambda}{4}(\phi^*\phi)^2, \qquad \frac{\partial V}{\partial\phi} = \frac{\mu^2}{2}\phi^* + \frac{\lambda}{2}\phi^*(\phi^*\phi) = 0 \qquad (5.28)$$

i.e. $\phi^* = 0$ (a local maximum) or $|\phi|^2 = -\mu^2/\lambda \equiv v^2$.

* Large perturbations about the minimum will not interest us. They can only occur is systems at very high energy density e.g. in very dense material such as neutron stars. Otherwise there are only minor effects due to barrier penetration.



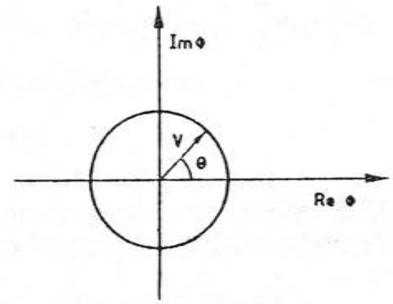



Fig. 5.2.

The minimum fixes $|\phi|^2$ to be $v^2$, i.e. the minimum is a circle in the $\phi$ plane.

Since $\theta$ (the phase angle of the complex field) is undetermined, we can fix the vacuum to be at any $\theta$ we wish. We will take $\theta = 0$. In order to make perturbations about this vacuum we now make the substitution

$$\phi = e^{i\xi/v}(v + \eta) , \qquad (5.29)$$

where $\xi$ represents perturbations in the $\theta$ direction and $\eta$ represents perturbations in the radial direction. For small $\xi$, $\eta$ this reduces to

$$\phi = v + \eta + i\xi , \qquad (5.30)$$

which is equivalent to doing perturbations about Re $\phi = v$ and Im $\phi = 0$. Substituting this into (5.27) we get

$$\mathcal{L} = \tfrac{1}{2}[(\partial_\mu\xi)^2 + (\partial_\mu\eta)^2] - \frac{\lambda v^2}{4}\eta^2 + \text{cubic and higher order terms} . \qquad (5.31)$$

This tells us that we have a particle $\eta$ with mass $\lambda v^2/2$. This mass is a consequence of trying to displace the $\eta$ against the restoring forces of the potential (the potential is like fig. 5.1(b) in the $\eta$ direction). The $\xi$ particle has no mass – it is known as a Goldstone boson; it corresponds to displacements around the minimum surface (i.e. around the circle in fig. 5.2) where there is no restoring force, since the potential is flat.

We can ask in this case, what happens if $\langle\phi\rangle$ takes different phase values in different places A and B. In between, we would like to keep the gradient of $\phi$ as small as possible in order to keep the energy as small as possible; but if A and B are far apart we can manage this because $\langle\phi\rangle$ can continuously vary as we go from A and B. As A and B get infinitely far apart, the gradients tend to zero and so the energy stored is zero. We can consider this changing of $\phi$ in the vacuum as a long wavelength excitation: as the wavelength tends to infinity (A and B tend to infinite separation) the energy tends to zero, so this excitation in the vacuum will correspond to a zero mass particle.

The need for such a massless particle appears to be general, and not dependent on our specific example. If the original Lagrangian has symmetry, this symmetry may be broken by the solution of minimum energy in the real world (the physical vacuum). But if the symmetry is represented by a continuous variable (like a rotation of phase) the "direction" of breaking can be slightly different in different places and waves due to perturbations in this direction must always be possible. In general, little energy is associated with long wave disturbances and we have (after quantising these perturbation waves) particles of necessarily zero mass – called Goldstone bosons. There are, however, cases where the variation of direction generates a current or charge density of some kind with which there are long range forces associated (i.e. $r^{-1}$ potentials). Then the large contribution of large volumes to the energy of interaction in the long wavelength waves, increases the energy $\omega$ of the long waves to a finite value; the quantised excitations are now of finite energy as the wavenumber $k \to 0$ and hence of finite mass. (This "mass" generation by long range force is familiar in solid state physics where density variations of neutral molecules give rise to phonons with a dispersion $\omega = C_s k$, but compressional oscillations of charges like electrons give rise to plasma waves with a dispersion $\omega = \sqrt{\omega p^2 + k^2}$ ($\omega p$ = a constant) due to the long range Coulomb interaction between the charge densities.)

Since later we shall want to use symmetry breaking to explain how mass terms arise and since zero mass Goldstone bosons are not found, we whall have to add long range interactions (of zero mass Yang–Mills fields) to give them mass by this mechanism, called the Higgs–Kibble mechanism [4,5,14]. If there is more than one way to vary $\phi$, while keeping the energy a minimum then there will be more than one Goldstone boson. The following example illustrates this. In SU(2), take an isovector $\boldsymbol{\phi}$; the minimum of the potential $V(\phi^+\phi)$ will occur for some non zero magnitude of $\phi$, but its direction is undetermined and we can take $\langle\phi\rangle = \begin{pmatrix}0\\0\\v\end{pmatrix}$. In order to look at small perturbations about the minimum we consider (by analogy with the previous example)

$$\boldsymbol{\phi} = \exp\left(\frac{i\xi_1 L_1}{v}\right)\exp\left(\frac{i\xi_2 L_2}{v}\right)\begin{pmatrix}0\\0\\v+\eta\end{pmatrix}, \tag{5.32}$$

where $L_1$ and $L_2$ are two of the three SU(2) generators. Now the minimum is on the surface of a sphere (cf. the circle in fig. 5.2) there are now two independent directions in which the potential is flat viz. the $\xi_1$ and $\xi_2$. The $\eta$ meson has a mass corresponding to radial displacements off the sphere, in which direction $V(\phi^+\phi)$ increases.



In general if the Lagrangian is invariant under a group G but the vacuum has a lower symmetry i.e. it is invariant under a group G′ where G′ < G. Then there will be massless bosons whose number $n$ = dim G – dim G′. In the SU(2) example above, the vacuum is invariant under U(1) so that the number of massless mesons is 3 – 1 = 2.

We give another example of this phenomenon which has more relevance to the physical world. Suppose the u and d quarks have zero mass. Then the Lagrangian

$$\mathcal{L} = \bar{\psi}_u i\partial\!\!\!/\psi_u + \bar{\psi}_d i\partial\!\!\!/\psi_d + \text{chirally invariant interaction terms}\,, \tag{5.33}$$

is invariant under the transformation

$$\psi \to e^{-ib\gamma_5}\psi, \qquad \bar{\psi} \to \bar{\psi}e^{-ib\gamma_5} \tag{5.34}$$

which is called a "chiral transformation" for arbitrary constant $b$. We can see this easily, as $\gamma_5$ anticommutes with the other $\gamma$ matrices. Note however that a mass term would not be invariant

$$\bar{\psi}m\psi \to \bar{\psi}me^{-2ib\gamma_5}\psi\,, \tag{5.35}$$

but a coupling to a vector potential via $\gamma_\mu$ would be invariant. Thus unless the symmetry is broken all the solutions of the Lagrangian must be chirally invariant and so, for example, the proton would have to be massless or parity doubled (because the chiral transformation changes the parity of a wavefunction). Experimentally the physical world does not have this symmetry. Therefore this symmetry must be broken and a Goldstone boson must exist if (5.33) is valid. No such massless particle exists but it would have to be a pseudoscalar meson, and therefore there is a temptation to associate it with the pion. Unfortunately the pion is not massless – but it is very light and people have therefore concluded the following. It might be that the u and d quarks have a small mass and therefore the chiral symmetry is only approximate; this would then possibly give a small mass to the pion. If this is the case, we can deduce a number of things about the pion couplings; these relations, such as the Goldberger–Treiman relation [e.g. 15], known as PCAC, are approximately verified experimentally. (We could expect another low mass boson coupled to $u\bar{u} + d\bar{d}$ i.e. a pseudoscalar but with isospin equal to zero. This is not found, people feel that the $\eta'$ is too heavy, and how this difficulty can be resolved has always been a fascinating problem (the problem of the ninth pseudoscalar boson in SU(3)). I am sorry to find that the length of this course is too short to permit me to discuss it along with many other interesting things I had hoped to discuss).

Let us return to vector fields. As we have so far discussed them, only the massless ones seem to make sense quantum mechanically. These have long range forces and we might therefore expect that the presence of these fields would eliminate the Goldstone boson when we break the symmetry. This is indeed the case as we will now show. Take a Lagrangian describing charged scalar particles interacting with photons

$$\mathcal{L}=\tfrac{1}{2}(\partial_\mu+\mathrm{i}A_\mu)\phi^*(\partial_\mu-\mathrm{i}A_\mu)\phi-\frac{\mu^2}{2}\phi^*\phi-\frac{\lambda}{4}(\phi^*\phi)^2-\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu}, \quad (5.36)$$

with $\mu^2 < 0$.

This is just the same as the Lagrangian (5.27) with electromagnetism added. We make the same substitution as before viz. $\phi = \mathrm{e}^{\mathrm{i}\xi/v}(v+\eta)$. Again we take the minimum of the potential to be $v$, giving

$$\mathcal{L}=\tfrac{1}{2}(\partial_\mu\xi)^2+\tfrac{1}{2}(\partial_\mu\eta)^2-\frac{\lambda v^2}{4}\eta^2+\tfrac{1}{2}v^2A^2-vA_\mu\partial_\mu\xi-\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu}. \quad (5.37)$$

It looks as if the $A_\mu$ field has acquired a mass (see the term $\frac{1}{2}v^2A^2$) but there is a peculiar term where $A_\mu$ is coupled to $\partial_\mu\xi$ but the coupling is really $\frac{1}{2}v^2(A_\mu-(\partial_\mu\xi/v))^2$ so we can remove this term by doing a gauge transformation

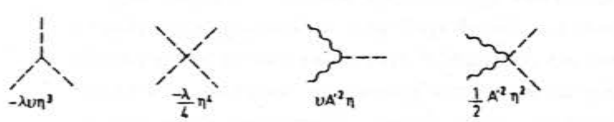
$$\phi' = \mathrm{e}^{-\mathrm{i}\xi/v}\phi, \qquad A'_\mu = A_\mu-(\partial_\mu\xi/v). \quad (5.38)$$

The Lagrangian then becomes

$$\mathcal{L}=\tfrac{1}{2}(\partial_\mu\eta)^2-\frac{\lambda v^2}{4}\eta^2+\tfrac{1}{2}v^2A'^2-\lambda v\eta^3-\frac{\lambda}{4}\eta^4$$

$$+\tfrac{1}{2}A'^2\eta^2+vA'^2\eta-\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} \quad (5.39)$$

($F_{\mu\nu}$ is invariant under this gauge transformation). In this Lagrangian we again have a massive field $A'_\mu$ but now the field $\xi$ has disappeared altogether and all the interactions are cubic or higher order. This phenomenon is called the Higgs mechanism. (There is no term like $A_\mu\partial_\mu\xi$.) These interactions may be represented by the following diagrams




(5.40)

where ------ is an $\eta$ propagator; ~~~~~~ is an $A'_\mu$ propagator.



One could say: haven't we lost a degree of freedom ($\xi$) between Lagrangians (5.36) and (5.39)? No, because the theory is no longer gauge invariant electrodynamics – the free vector particle $A'_\mu$ now has three polarisation states since it has a mass, whereas the massless $A_\mu$ only had two polarisation states. There are four dynamical degrees of freedom in each case. We lost the gauge invariance when we made the explicit gauge transformation to eliminate $\xi$ and a gauge transformation on (5.39) will not leave it invariant; in particular, it will bring back a term of the form $A_\mu\partial_\mu\xi$ where $\xi$ is the gauge parameter.

We will now give a physical example of this phenomenon. The example is superconductivity and it is discussed in detail in [16]; here we shall merely give an outline. Superconductivity has the same properties as described above* except that it is non-relativistic. In the case of a metal at low temperatures, the electrons form (Cooper) pairs of opposite spin in such a way that these pairs act as bosons. Let $\psi$ be the wave function for one of these bosons; it satisfies the Schroedinger equation. We consider the case of $\psi$ interacting with an electromagnetic field

$$\frac{1}{2m}(\mathrm{i}\underline{\nabla}-\mathrm{e}\underline{A})^2\psi = E\psi, \quad (5.41)$$

where $E$ is a constant and $e$ is the charge of a pair of electrons. Since the $\psi$ is a boson, many pairs of electrons can be in the same state. The electromagnetic current is

$$j=\frac{1}{2m}[[(\mathrm{i}\underline{\nabla}-\mathrm{e}\underline{A})\psi]^*\psi+[\psi^*(\mathrm{i}\underline{\nabla}-\mathrm{e}\underline{A})\psi]]. \quad (5.42)$$

Suppose that in the absence of $\underline{A}$ there is no current but that all particles are in the same state because of the Bose condensation. This $j$ represents not only the probability current of one particle, but when multiplied by $e$ and $N/2$ (the number of particles) represents the physical electric current of the bosons. When very many bosons are in the same state, the wave function acquires a real physical significance, just as the photon wave function becomes the physically real (gauge transformations excepted) $\underline{A}(x,t)$ of classical mechanics and electromagnetism (Maxwell theory) when there are sufficient numbers of photons in the same state to make a "real" light wave. Turn on a very small field $\underline{A}$. To a first approximation, in many cases (those which yield superconductivity) the wavefunction is unchanged because so many interacting bosons are in the same state; so that the current we get is

* The Higgs Lagrangian (5.36) in the static case is identical to the Ginzburg–Landau free energy in the theory of type II superconductors [17].

$$j = \underline{A}\,\psi^*\psi/m\,. \tag{5.43}$$

But $\psi^*\psi$ is simply the density of bosons; call this $N/2$ ($N$ is the density of the electrons). The current is

$$\underline{j}_{\text{super}} = \frac{N}{2}\left(\frac{e^2\underline{A}}{m}\right) \tag{5.44}$$

i.e. the current in a superconductor is proportional to the vector potential. The Maxwell equation

$$\nabla^2\underline{A} = \underline{j}\,, \tag{5.45}$$

becomes

$$\nabla^2\underline{A} = \underline{j}_{\text{external}} + \underline{j}_{\text{superconducting}}\,, \tag{5.46}$$

i.e.

$$(\nabla^2 - \Lambda)\underline{A} = \underline{j}_{\text{external}}\,, \tag{5.47}$$

where $\Lambda = Ne^2/2m$ and is called the London constant. This is the equation which describes the behaviour of a vector particle with a mass. The theory has spontaneously acquired a mass exactly as in the Higgs case. Notice that we have also lost the gauge invariance as we did in the Higgs case; we have chosen the gauge where the equations describing the physical properties of the theory are in simplest form.

Recall that when we considered the geometrical significance of gauge invariance in sect. 3 we saw how the gauge information could be carried from place to place by a particle. In this problem we have merely chosen the gauge which is carried by the electrons and it is clear that this is the natural gauge for this problem; this is why the equations appear simpler in this gauge.

We shall now look at spontaneous symmetry breakdown in Yang–Mills theories. We look at two cases

(i) Isospinor scalar particles. We construct the same $V(\phi)$ as we had earlier

$$V(\boldsymbol{\phi}) = \frac{\mu^2}{2}\,\phi_i^+\phi_i + \frac{\lambda}{4}\,(\phi_i^+\phi_i)^2\,, \tag{5.48}$$

where $i = 1, 2$. Let us assume that at some point in space

$$\boldsymbol{\phi} = v\begin{pmatrix}1\\0\end{pmatrix}. \tag{5.49}$$

If $\phi$ is pointing in the three direction of the three dimensional space of SU(2), there will no longer be gauge independence in the 1 and 2 directions since if



we rotate about the 1 or 2 directions $\phi$ changes, so we expect that two components of the $A_\mu$ field pick up mass. But what about the 3 direction? When we rotate about the 3 direction, a spinor gets multiplied by a phase i.e.

$$\begin{pmatrix}1\\0\end{pmatrix} \to \begin{pmatrix}e^{i\theta/2}\\0\end{pmatrix}, \tag{5.50}$$

for a rotation through $\theta$. So we do not have any gauge freedom left and therefore all three components of $A_\mu$ pick up a mass.

Mathematically

$$\mathcal{L} = -\frac{1}{4g^2}E_{\mu\nu}E_{\mu\nu} + \tfrac{1}{2}(\partial_\mu + i\boldsymbol{\tau}\cdot A)\boldsymbol{\phi}^+(\partial_\mu - i\boldsymbol{\tau}\cdot A)\boldsymbol{\phi} - V(\boldsymbol{\phi})\,. \tag{5.51}$$

Put $\phi = \binom{1}{0}\,v$ and look at the mass term for the $A$'s. This is

$$\tfrac{1}{2}v^2(1 \quad 0)(\boldsymbol{\tau}\cdot A)(\boldsymbol{\tau}\cdot A)\begin{pmatrix}1\\0\end{pmatrix} = \frac{v^2}{8}[A_1^2 + A_2^2 + A_3^2]\,. \tag{5.52}$$

So we see that all three components of $A$ have acquired the same mass.

(ii) Isovector scalar particles

$$\boldsymbol{\phi} = \begin{pmatrix}\phi_1\\ \phi_2\\ \phi_3\end{pmatrix}.$$

The potential has the same form as (5.48) with $i = 1, 2, 3$. Let the vacuum expectation value of $\phi$, $\langle\phi\rangle = \begin{pmatrix}0\\0\\1\end{pmatrix}v$, define the 3 direction in isospace. Carrying this $\phi$ particle around tells us where the 3 directions is everywhere. Hence, we lose the freedom to rotate the $\phi$ field about the 1 and 2 directions and therefore $A^1$ and $A^2$ will become massive – we have broken the gauge invariance in the 1 and 2 directions. But if we rotate by $\theta$ about the 3 direction the 1 component is multiplied by $e^{i\theta}$, the 2 component is multiplied by $e^{-i\theta}$, and the 3 component is left unchanged i.e. we still have a freedom of gauge rotation about the 3 direction. $A^3$ will still have zero mass.

Mathematically,

$$\mathcal{L} = \frac{1}{4g^2}E_{\mu\nu}E_{\mu\nu} + \tfrac{1}{2}[i(\partial_\mu + A_\mu\times)\boldsymbol{\phi}]^+[i(\partial_\mu + A_\mu\times)\boldsymbol{\phi}] - V(\boldsymbol{\phi}) \tag{5.53}$$

Substituting $\phi = \begin{pmatrix}0\\0\\1\end{pmatrix}v$ we get the mass term for $A$:

$$\frac{v^2}{2}[(A^1)^2 + (A^2)^2] \tag{5.54}$$

i.e. $A^1$ and $A^2$ have the same mass and $A^3$ is massless, as expected. One might ask – why do we use scalar particles to break the symmetry? They are of course the simplest to write down. But this "superconductivity" effect can be generated in many ways (as indeed in real electrodynamics where pairs of fermions do it), and it is possible that the symmetry breaking if it occurs, in say, weak interactions or other places in physics, may have a more complex mechanism than the Higgs scalar method. But to-day we have a severe restriction on the meaningful theories we can write down i.e. that they are relativistic, quantum mechanical, and renormalisable. If all these restrictions are imposed it looks as if only the Higgs method can be formulated at present.

## 6. Quantisation

### *6.1. Philosophy*

Before we commence a detailed study of the quantisation of a Yang–Mills fields, we shall describe the various approaches to quantising a classical theory. Fig. 6.1 illustrate a general schema for quantising particle mechanics.

We begin with a Lagrangian from which we can deduce a classical Hamiltonian, written in terms of $q(t)$ and its conjugate momentum $p(t)$, defined by $p(t) = \mathrm{d}\mathcal{L}/\mathrm{d}\dot{q}(t)$. There are two routes by which we can quantise the theory.

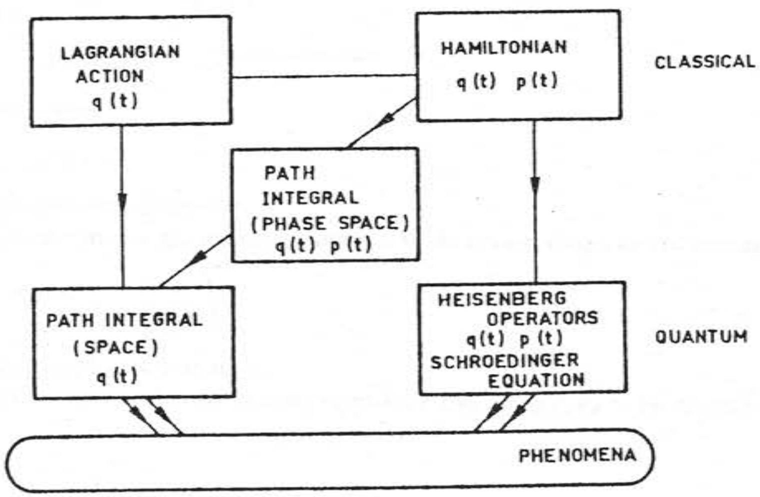



Fig. 6.1.



From the Hamiltonian we can define operators associated with $p(t)$ and $q(t)$ and hence obtain either the Heisenberg or Schroedinger pictures; from these we can deduce results which predict the behaviour of phenomena in the physical world. There may be a problem with operator ordering when we go from the classical to the quantum theory. The alternative approach is to start with the Lagrangian and introduce a path integral which associates a certain amplitude with each trajectory in space [18]; this enables us to proceed directly to calculate the consequences of the quantum theory. (There exists an alternative path integral method, not much used these days, devised by DeWitt–Morette [19] and Garrod [20] in which one can construct a path integral in phase space directly from the Hamiltonian. This resolves many questions of the order of operators in the Hamiltonian. From it the ordinary path integral is easily obtained.) All methods lead to the same consequence physically, but each method has its advantages and disadvantages and we choose which ever is the more convenient for the problem at hand e.g. spin $\frac{1}{2}$ is very awkward to handle in path integral approach.

Consider now the situation when we try to quantise a field theory (fig. 6.2).

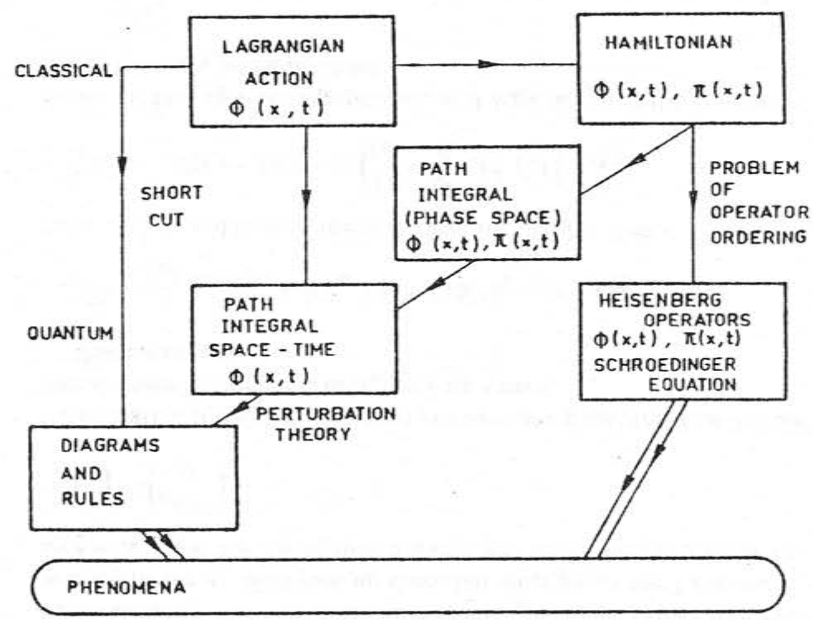



Fig. 6.2.

The classical variables $q(t)$ and $p(t)$ are replaced by $\phi(x, t)$ and its conjugate momentum $\pi(x, t)$ defined by $\pi(x, t) = \partial\mathcal{L}/\partial\dot{\phi}(x, t)$. The situation is similar to that in quantising a particle theory and historically people like Heisenberg and Dirac [e.g. 21] followed the Hamiltonian approach. This leads to $\phi(x, t)$ and $\pi(x, t)$ being operators and again we have a problem with operator ordering in the more complicated field theories. There is also a Schroedinger approach which people don't use much these days in which the vacuum has a wave function in terms of the co-ordinates i.e. an explicit functional of the field function $\phi(x)$. Alternatively we can proceed by the path integral method; the path is now a function of $\phi(x, t)$. Because of difficulties in calculating with the theory we usually proceed by perturbative methods. To do this we normally use diagrams and rules, all of which can easily be deduced from the path integral formulation. We can also get these rules from a Heisenberg approach, but it is more difficult. Nowadays we know of an obvious and simple-minded short cut to get straight from the Lagrangian to the diagram rules. Some people who are not sufficiently acquainted with the theory, think that the rules are all there is, and then say that the theory is only defined by a perturbation expansion. It is true that we can only calculate things by perturbation theory, but this may be only a limitation of the era and in any case there are certain things that we can deduce from the Lagrangian without using perturbation theory. (For non-relativistic field theories, such as those that arise in solid state physics, we are not at all limited to perturbation theory, and many methods and solutions for large or intermediate couplings are known.)

In a field theory there is a problem of renormalisability, because, when we calculate diagrams with closed loops we get infinite answers. When we go through the Heisenberg approach, the theory is not manifestly Lorentz invariant; but in the path integral approach it is, so for this reason the latter approach is to be preferred when we attempt to renormalise the theory. The difficulty is to keep the renormalisation process Lorentz invariant when the form of the equations is not manifestly invariant. There is difficulty in the path integral approach, however, and that is concerned with the inclusion of fermions. The path integral method doesn't work in this case: but when people went through the operator approach they found that there were only differences in signs when fermions were introduced, and so minor were these differences that they forced the path integral formalism to work by introducing Grassmann algebras.

In attempting to quantise Yang–Mills theory, Schwinger [22], after a lot of hard work, found the Hamiltonian from the Lagrangian, but an attempt to proceed with the Hamiltonian approach ran into serious difficulties with the ordering of operators and progress in this direction ceased. I took the short cut



(in fig. 6.2) [23] and found that there are certain complications in the diagrams at the one loop level; but I got round these by introducing a contribution from a fictitious particle. The correct rules for this particle were first worked out by de Witt [24] and subsequently understood in a more general way by Fadde'ev and Popov [26]. However I could only do it for one closed loop; if there were two or more closed loops I didn't know what to do – I sort of half understood it. Fadde'ev and Popov straightened this problem out by going via the path integral method and discovered that to all orders we have to add a closed loop of scalar particle with Fermi statistics for every closed loop with a Yang–Mills vector meson, they also went via the Morett path integral approach using the Hamiltonian derived by Schwinger. The problem of renormalisability was solved by 't Hooft [25] using his method of dimensional regularisation.

Yang–Mills theory is often presented as being complicated, but now that all this work is done and proofs proved, it is really not much more complicated than QED apart from the more complex algebra which is involved; and a bit more care is needed with the gauge invariance. We just add the contribution of the ghost and regulate by the dimensional regularisation scheme.

It would be sensible now to give the correct and complete theory, say as outlined in Abers and Lee [10] and many students might prefer this. But to get a clear feeling for the need for the Fadde'ev–Popov ghost, we can contrast the theory with QED by asking what happens if we just plough along and make diagram rules by direct analogy with QED. This is a sort of "damn the torpedoes, full speed ahead" approach. We will discover that we are hit by a torpedo, but let's try it anyhow.

We will attempt to calculate with the theory by first deriving the diagram rules directly from the Lagrangian.

## *6.2. Derivation of the rules for diagrams*

We shall assume that the reader is familiar with the methods of deriving the rules from the Lagrangian and so we will merely give an outline of the derivation as we go along. Consider the following model Lagrangian for spin $\frac{1}{2}$ isospinors $\psi$ and spin 0 isovectors $\boldsymbol{\phi}$ interacting with Yang–Mills fields $A_\mu$

$$\mathcal{L} = -\tfrac{1}{4}E_{\mu\nu}E_{\mu\nu} + \bar{\psi}(\mathrm{i}\not{\partial} + g\not{A}\cdot\boldsymbol{\Psi})\psi - m\bar{\psi}\psi$$

$$+\tfrac{1}{2}[\mathrm{i}(\partial_\mu + gA_\mu\times)\boldsymbol{\phi}]^+[\mathrm{i}(\partial_\mu + gA_\mu\times)\boldsymbol{\phi}] - \frac{M^2}{2}\boldsymbol{\phi}^+\boldsymbol{\phi}, \tag{6.1}$$

where we have rescaled $A_\mu$ to $gA_\mu$ and so now

$$E_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + gA_\mu \times A_\nu . \tag{6.2}$$

We separate the Lagrangian into terms of second order and terms of third and higher orders in the fields. The second order terms give the propagators and the higher order terms give the interactions:

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}\cdot F_{\mu\nu} + \bar{\psi}(\mathrm{i}\gamma_\mu\partial_\mu - m)\psi + \tfrac{1}{2}[(\partial_\mu\phi)^+(\partial_\mu\phi) - M^2\phi^+\phi] \qquad \text{2nd order}$$

$$+ g\bar{\psi}\gamma_\mu A_\mu\cdot\tau\psi - g(A_\mu\times A_\nu)\cdot\partial_\mu A_\nu + \tfrac{1}{2}g[\partial_\mu\phi^+\cdot(A_\mu\times\phi) + \partial_\mu\phi\cdot(A_\mu\times\phi^+)] \qquad \text{3rd order}$$

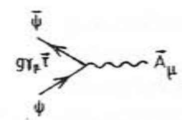
$$+ \tfrac{1}{2}g^2(A_\mu\times\phi^+)\cdot(A_\mu\times\phi) - \tfrac{1}{4}g^2(A_\mu\times A_\nu)\cdot(A_\mu\times A_\nu) \qquad \text{4th order} \tag{6.3}$$

where we have retained $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ to pull out second order terms in $A_\mu$ which will give the vector field propagator.

As there is some problem in deriving the propagator for the $A_\mu$ field, we will first derive the diagram rules for the interaction terms. Consider the term

$$g\bar{\psi}\gamma_\mu A_\mu\cdot\tau\psi\,, \tag{6.4}$$

this corresponds to the following interaction — $\psi$ coming in, absorbing or emitting an $A_\mu$ and going out as a $\bar{\psi}$, viz.

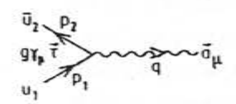


(6.5)

In co-ordinate space we take

$$\psi = u_1\,\mathrm{e}^{-\mathrm{i}p_1x}, \qquad \bar{\psi} = \bar{u}_2\,\mathrm{e}^{+\mathrm{i}p_2x}, \qquad A_\mu = a_\mu\,\mathrm{e}^{+\mathrm{i}qx}\,, \tag{6.6}$$

i.e. free waves. Going to momentum space the diagram becomes



(6.7)

where the momenta carried by the particles are indicated on the diagram. We can now read off the vertex; it is

$$g\bar{u}_2(\gamma_\mu\,\tau\cdot a_\mu)u_1\,. \tag{6.8}$$

This, except for the presence of the $\tau$ matrix, is identical to QED. In fact, as in QED, we do not need actually to write the $u_1, \bar{u}_2$ spinors for virtual fermions, but just string the Dirac (and $\tau$) matrices together in a product in order, as we come to them following a fermion line.



We now consider, in a similar fashion, the third order (in the fields) term for the scalar field $\phi$ interacting with the Yang–Mills field $A_\mu$

$$+\tfrac{1}{2}g\{\partial_\mu\phi^+\cdot(A_\mu\times\phi) + \partial_\mu\phi\cdot(A_\mu\times\phi^+)\}. \tag{6.9}$$

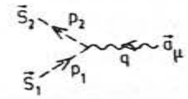
Putting in, as before, the free fields

$$\phi = S_1\,\mathrm{e}^{-\mathrm{i}p_1x}, \qquad \phi^+ = S_2\,\mathrm{e}^{+\mathrm{i}p_2x}, \qquad A_\mu = a_\mu\,\mathrm{e}^{-\mathrm{i}qx}\,, \tag{6.10}$$

we get

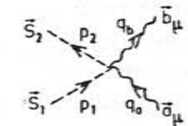
$$+\tfrac{1}{2}g\{+\mathrm{i}p_{2\mu}S_2\cdot(a_\mu\times S_1) - \mathrm{i}p_{1\mu}S_1\cdot(a_\mu\times S_2)\}\exp\{-\mathrm{i}(p_1 - p_2 + q)x\}, \tag{6.11}$$

then going to momentum space we have



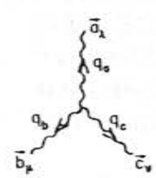
$$\mathrm{i}g(p_1 + p_2)_\mu[S_2\cdot(a_\mu\times S_1)] \tag{6.12}$$

Similarly, the term $\tfrac{1}{2}g^2(A_\mu\times\phi^+)\cdot(A_\mu\times\phi)$ gives



$$g^2[(b_\mu\times S_2)\cdot(a_\mu\times S_1) + (a_\mu\times S_2)\cdot(b_\mu\times S_1)]\,. \tag{6.13}$$

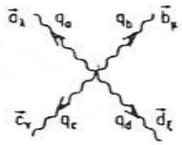
So far, these interaction are similar to (apart from the isospin labels) those in QED. But in Yang–Mills theory there are two extra vertices which have no QED analogue (they arise because the $A_\mu$ field itself carries "charge"). The term $-g(A_\mu\times A_\nu)\cdot\partial_\mu A_\nu$ gives, in momentum space, $-g(a_\mu\times b_\nu)\cdot c_\nu q_{c_\mu}$ when we substitute $a\,\mathrm{e}^{\mathrm{i}q_ax}$ for the first, $b_\mu\,\mathrm{e}^{\mathrm{i}q_bx}$ for the second, and $c_\nu\mathrm{e}^{\mathrm{i}q_cx}$ for the third with momenta $q_a, q_b, q_c$. Hence altogether it gives, taking account of permutations and with a bit of re-arranging of the dots and crosses:



$$-g\{(q_a - q_c)_\mu b_\mu\cdot(c_\nu\times a_\nu) + (q_c - q_b)_\mu a_\mu\cdot(b_\nu\times c_\nu) + (q_b - q_a)_\mu c_\mu\cdot(a_\nu\times b_\nu)\}. \tag{6.14}$$

There are six terms altogether since each of the three $A_\mu$'s can be either $a_\mu$, $b_\mu$ or $c_\mu$ in every possibly way.

The term $-\frac{1}{4}g^2(A_\mu \times A_\nu)\cdot(A_\mu \times A_\nu)$ gives



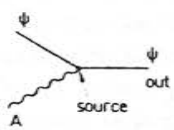

$$-\frac{g^2}{4}\{(a_\mu \times b_\nu)\cdot(c_\mu \times d_\nu) + \text{three other symmetric combinations}\}\,, \tag{6.15}$$

which comes to a total of $-g^2(a_\mu \times b_\nu)\cdot(c_\mu \times d_\nu)$ by a bit of rearranging.

We must finally derive the propagators for the theory from the quadratic terms in (6.3). The propagators for $\psi$ and $\phi$ are easy to get. The equation of motion for $\psi$ is (obtained from the Lagrangian by varying with respect to $\bar{\psi}$)

$$(\mathrm{i}\not{\partial} - m)\psi = \not{J}\,, \tag{6.16}$$

where $J$ is some source for $\psi$ (whose exact form is irrelevant, e.g. some non-linear combination of other fields like $A\cdot\tau\,\psi$, that in the diagrams generate a source away from which a particle $\psi$ is to propagate)




So symbolically

$$\psi = \frac{1}{\mathrm{i}\not{\partial} - m}\not{J}\,. \tag{6.17}$$

In momentum space this is

$$\psi_p = \frac{1}{\not{p} - m}\not{J}_p \tag{6.18}$$

This gives the $\psi$ which enters into a source term for another interaction, and so each virtual $\psi$ line brings in a factor $(\not{p} - m)^{-1}$. This then defines the propagator


$$\frac{\delta_{ij}}{\not{p} - m}\,. \tag{6.19}$$

The $i$ and $j$ are isospin indices. The $\delta_{ij}$ is present since $\psi$ must couple to the same isospin at both ends.



For the $\phi$ field, the equation of motion is

$$(\Box + M^2)\phi = -J'\,. \tag{6.20}$$

Therefore

$$\phi_p = \frac{1}{p^2 - M^2}J'_p\,, \tag{6.21}$$

where we have again transformed to momentum spa... ...is gives the $\phi$ propagator




$$\frac{\delta_{ab}}{p^2 - M^2}\,. \tag{6.22}$$

where the $a$ and $b$ are isospin indices.

Finally consider the $A_\mu$ propagator; the equation of motion is

$$\partial_\mu(\partial_\mu A_\nu - \partial_\nu A_\mu) = -J'_\nu\,, \tag{6.23}$$

i.e.

$$\Box A_\nu - \partial_\nu(\partial_\mu A_\mu) = -J'_\nu \tag{6.24}$$

where $J'_\nu$ is the total current – matter plus contributions from the $A_\mu$ itself due to the third and higher order terms in the Lagrangian. We need to solve this equation to get the propagator. However in general we cannot solve it because by itself it is meaningless unless the divergence of $J'_\nu$ is zero, for the dive gence of the left side is identically zero. We conclude therefore that

$$\partial_\nu J'_\nu = 0\,,$$

i.e.

$$\partial_\nu\{J_\nu + \partial_\mu(A_\mu \times A_\nu) + A_\mu \times E_{\mu\nu}\} = 0\,. \tag{6.25}$$

This equation can indeed be verified using the relation (2.42).

When we attempt to obtain the propagator in QED, we do so by choosing

$$\partial_\mu A_\mu = 0\,, \tag{6.26}$$

as we may do due to gauge invariance. The equation of motion then becomes simply

$$\Box A_\mu = -J_\mu\,, \tag{6.28}$$

and we can solve this to obtain the propagator viz.

$$A_\mu = \frac{\delta_{\mu\nu}}{p^2} J_\nu \,, \tag{6.29}$$

so the propagator is

$$\delta_{\mu\nu}/p^2 \,. \tag{6.30}$$

It turns out that when we use this and calculate diagrams in QED, the current we get from the diagrams is automatically conserved so that everything is alright, as it must be since

$$\partial_\mu \Box A_\mu = -\partial_\mu J_\mu \,, \tag{6.31}$$

from (6.28), and both sides are separately zero; so the gauge choice is self consistent. This is equivalent to replacing the action $-\frac{1}{4}\int F_{\mu} F_{\mu\nu} d^4x$ by $-\frac{1}{2}\int(\partial_\nu A_\mu)(\partial_\nu A_\mu) d^4x$, or in other words to adding an extra term $+\frac{1}{2}\int(\partial_\mu A_\mu)^2 d^4x$ to the action, so the Lagrangian becomes simply $-\frac{1}{4}F_{\mu\nu}F_{\mu\nu} + \frac{1}{2}(\partial_\mu A_\mu)(\partial_\nu A_\nu)$. Take a flying guess and try to do the same thing for Yang–Mills – an excellent method of doing physics. The purpose of physics is to find out what's true, not to find out what you can prove. If you allow yourself the liberty of knowing things with different degrees of certainty, then you can know a lot more physics than if you have to prove everything. So in (6.24) we suppose that we can choose the gauge

$$\partial_\mu A_\mu = 0 \,, \tag{6.32}$$

and we then get the following for the equation of motion

$$\Box A_\nu = -J'_\nu \,. \tag{6.33}$$

We can invert this equation to get

$$A_\nu = \frac{\delta_{\nu\mu}}{\Box}(-J'_\mu) \tag{6.34}$$

which in momentum space is

$$(A_p)_\nu = \frac{\delta_{\nu\mu}}{p^2}(J'_p)_\mu \,. \tag{6.35}$$

This defines the propagator


$$\frac{\delta_{\mu\nu}\delta_{ab}}{p^2} \,. \tag{6.36}$$

Notice that if we take the divergence of both sides of (6.33), both sides are



separately zero, the left side by the gauge condition (6.32) and the right side by (6.25) so we appear to be self consistent. We have all the rules and can calculate processes and see whether we run up against any problems, such as gauge non invariance of a physical process. We will, indeed, run into problems when we consider Yang–Mills loops and shall see that we need to introduce something extra (the ghost particle) to cure them.

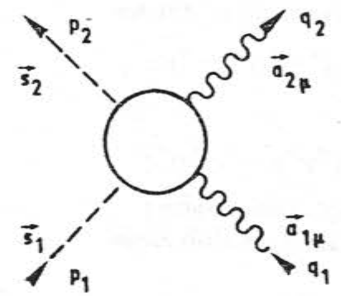
### 6.3. *Explicit calculations of physical processes in Yang–Mills theory*

As an application of the rules that we have derived, we shall now calculate several physical processes. We begin with Compton scattering for $I = 1$ scalon.

The process under consideration here is Compton scattering of a gluon $A_\mu$ and an isovector scalar particle $\phi$ (scalon) using the Lagrangian (6.1)

$$\text{gluon} + \text{scalon} \to \text{gluon} + \text{scalon} \,. \tag{6.37}$$

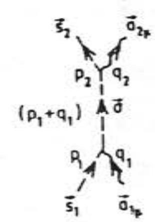
We define the momenta, polarisation vectors and isospin labels as follows




(6.38)

The momenta are $p_1, p_2$ and $q_1, q_2$, the isospin labels of the scalons are $S_1$ and $S_2$ and the polarisation vectors and isospin labels of the gluons are $a_{1\mu}$ and $a_{2\mu}$. To second order in the coupling constant there are four diagrams which contribute to this process, and we will calculate them in turn

(i)




(6.39)

where we have temporarily put an isospin index $\boldsymbol{\sigma}$ on the intermediate scalon to make it easier to apply our rules. The amplitude for this process is (we must sum over all isospin directions of $\boldsymbol{\sigma}$ as all will contribute to the diagram)

$$\sum_{\substack{\text{isospin}\\ \text{directions}}} g^2(2p_2+q_2)_\nu S_2\cdot(a_{2\nu}\times\boldsymbol{\sigma})\,\frac{1}{(p_1+q_1)^2-m^2}$$

$$\times\,\boldsymbol{\sigma}\cdot(a_{1\mu}\times S_1)(2p_1+q_1)_\mu\,, \tag{6.40}$$

Re-writing $S_2\cdot(a_{2\nu}\times\boldsymbol{\sigma})$ as $(S_2\times a_{2\nu})\cdot\boldsymbol{\sigma}$ and using the formula

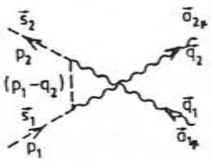
$$\sum_{\sigma}(A\cdot\boldsymbol{\sigma})(\boldsymbol{\sigma}\cdot B)=A\cdot B\,, \tag{6.41}$$

we get for the contribution from this diagram

$$g^2\,\frac{(2p_2+q_2)_\nu(S_2\times a_{2\nu})\cdot(a_{1\mu}\times S_1)(2p_1+q_1)_\mu}{(p_1+q_1)^2-m^2}. \tag{6.42}$$

The second diagram is the same as (i) with the two gluons interchanged.

(ii)




(6.43)

The amplitude for this diagram can be obtained immediately from (i) if we notice that making the substitutions

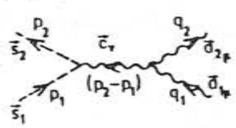
$$q_1\leftrightarrow -q_2,\qquad a_{1\mu}\leftrightarrow a_{2\mu}\,, \tag{6.44}$$

in (i) gives (ii). Therefore the contribution from (ii) is

$$g^2\,\frac{(2p_2-q_1)_\nu(S_2\times a_{1\nu})\cdot(a_{2\mu}\times S_1)(2p_1-q_2)_\mu}{(p_1-q_2)^2-m^2}. \tag{6.45}$$

The third diagram is

(iii)




(6.46)



$$\text{amplitude}=-\frac{(S_2\times S_1)\cdot c_\lambda(p_1+p_2)_\lambda}{(p_1-p_2)^2}\{(q_1-p_1+p_2)_\nu a_{2\nu}\cdot(a_{1\mu}\times c_\mu)$$

$$+(-q_2-q_1)_\nu c_\nu\cdot(a_{2\mu}\times a_{1\mu})+(p_1-p_2+q_2)_\nu a_{1\nu}\cdot(c_\mu\times a_{2\mu})\}\,,$$

where $c_\nu$ is the polarisation vector of the intermediate gluon, and therefore we must sum over its directions in space and isospace using the formula

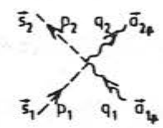
$$\sum_{c_\nu}(A\cdot c_\nu)(c_\mu\cdot B)=A\cdot B\delta_{\mu\nu}\,. \tag{6.47}$$

The amplitude for this diagram is then

$$-g^2\,\frac{(S_2\times S_1)}{(p_1-p_2)^2}\{(a_{2\nu}\times a_{1\mu})[(p_1+p_2)_\mu(2q_1-q_2)_\nu$$

$$+(p_1+p_2)_\nu(2q_2-q_1)_\mu]-(a_{2\mu}\times a_{1\mu})(p_1+p_2)_\lambda(q_1+q_2)_\lambda\}. \tag{6.48}$$

Finally we have the diagram




(6.49)

The amplitude is

$$g^2\{(S_2\times a_{2\mu})\cdot(S_1\times a_{1\mu})+(S_2\times a_{1\mu})\cdot(S_1\times a_{2\mu})\}. \tag{6.50}$$

Adding the contributions from each of the diagrams, rearranging dot and cross vector products, putting the scalons on mass shell and using momentum conservation (but keeping the gluons off mass shell for the moment) we get the total amplitude for the process to be

$$(S_2\times a_{2\nu})\cdot(S_1\times a_{1\mu})\left\{\frac{(2p_2+q_2)_\nu(2p_1+q_1)_\mu}{2p_1q_1+q_1^2}\right.$$

$$-\frac{(2q_1-q_2)_\nu(p_1+p_2)_\mu+(p_1+p_2)_\nu(2q_2-q_1)_\mu}{(q_1-q_2)^2}$$

$$\left.+\frac{(p_1+p_2)\cdot(q_1+q_2)+(q_1-q_2)^2}{(q_1-q_2)^2}\,\delta_{\mu\nu}\right\} \tag{6.51}$$

*equation continued on next page.*

$$+(S_1 \times a_{2\nu})\cdot(S_2 \times a_{1\mu})\left\{-\frac{(2p_1-q_2)_\nu(2p_2-q_1)_\mu}{-2p_2q_2+q_2^2}\right.$$

$$+\frac{(2q_1-q_2)_\nu(p_1+p_2)_\mu+(p_1+p_2)_\nu(2q_2-q_1)_\mu}{(q_1-q_2)^2}$$

$$\left.+\frac{-(p_1+p_2)\cdot(q_1+q_2)+(q_1-q_2)^2}{(q_1-q_2)^2}\,\delta_{\mu\nu}\right\}. \qquad (6.51, \text{cont'd})$$

So what? What do we do with the answer? There are two possibilities – we can use it to calculate the Compton scattering – maybe we are interested in this; or we can use it to calculate the annihilation of a pair of scalons into a pair of gluons (see below).

In order to get the Compton scattering of real gluons we must put them on their mass shell i.e. $q_1^2=q_2^2=0$ but this alone is not sufficient, as we can see as follows. Consider the equation of motion for the gluons in the absence of any sources (free gluons):

$$\partial_\mu(\partial_\mu A_\nu - \partial_\nu A_\mu) = 0 \qquad (6.52)$$

In momentum space, this becomes

$$q_\mu(q_\mu a_\nu - q_\nu a_\mu) = 0 \qquad (6.53)$$

i.e.

$$q^2 a_\nu - (q_\mu a_\mu) q_\nu = 0\,. \qquad (6.54)$$

There are two solutions to this equation: either (a) $q^2=0$ and we must have $q_\mu a_\mu = 0$. I.e. for a Yang–Mills particle on its mass shell, the polarisation vector is perpendicular to the momentum; or (b) $q^2 \neq 0$; then $a_\nu = (q_\mu a_\mu/q^2)q_\nu$, i.e.

$$a_\nu = \alpha q_\nu\,. \qquad (6.55)$$

This corresponds in co-ordinate space to $A_\nu = \partial_\nu \chi$ which is a pure gradient. So we would perhaps expect that this field could be removed by an infinitesimal gauge transformation. Let us try this; to gauge it away we would have to have

$$0 = \partial_\nu \chi + \partial_\nu \alpha + (\partial_\nu \chi \times \alpha)\,.$$

Since $\partial_\nu \chi \times \alpha$ is higher order in the field, it can be neglected since we are working with a free wave and hence $a_\nu$ must be small (otherwise there would be self coupling terms on the right hand side of 6.52 which we neglected in order to get the free wave solution, since they are of second and higher order in the fields). We can therefore take $\alpha$ to be $-\chi$ and gauge the field away.



Case (b) cannot produce any physics. If we calculate an amplitude for a physical process first with $a_\mu$, and then with $a'_\mu$, where $a'_\mu$ is related to $a_\mu$ by an infinitesimal gauge transformation $a'_\mu = a_\mu + \partial_\mu \alpha$, there should be no difference in the answers since the theory is gauge invariant. The effect of $a_\mu$ and $a'_\mu$ is linear and this means that if we were to calculate with $\partial_\mu \alpha$ the amplitude must be zero.

So we can test for the correctness of the solution to any physical process by putting the polarisation vector $a_\mu$ of an external gluon equal to $\alpha q_\mu$ where $q_\mu$ is its momentum, and checking that this gives zero. Note that we must carry out this test for a whole physical process; we should not expect that a single diagram by itself is gauge invariant as it is only part of an answer and by itself represents no physics.

In the Compton effect above, we gauge the incoming gluon, putting

$$a_{1\mu} = \alpha q_{1\mu} \qquad (6.56)$$

Do we get zero? We make the substitution, and of course we don't get zero – because we made a couple of mistakes; however fiddling around a bit more and correcting the mistakes, the Compton amplitude (6.51) becomes

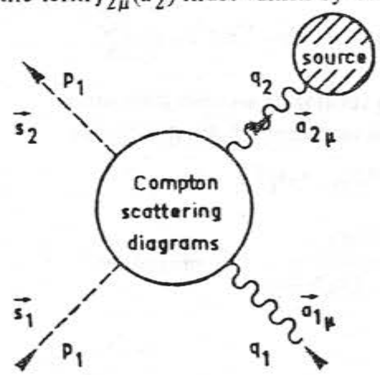
$$-g^2(S_1 \times \alpha)\cdot(S_2 \times j_{2\mu}(a_2))\left\{\frac{(2p_1+q_1)_\mu}{(q_1-q_2)^2}\right\}$$

$$-g^2(S_2 \times \alpha)\cdot(S_1 \times j_{2\mu}(a_2))\left\{\frac{(-2p_2+q_1)_\mu}{(q_1-q_2)^2}\right\}$$

where

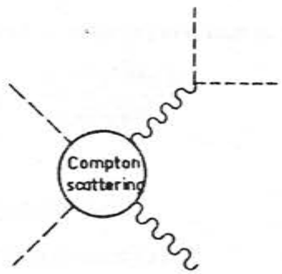
$$j_{2\mu}(a_2) = q_2^2 a_{2\mu} - q_{2\mu}(q_{2\nu}a_{1\nu})\,. \qquad (6.57)$$

Is this zero? Yes, because $a_{2\mu}$ is supposed to represent a free sourceless gluon, so the term $j_{2\mu}(a_2)$ must vanish by 6.54. If $a_{2\mu}$ is not sourceless



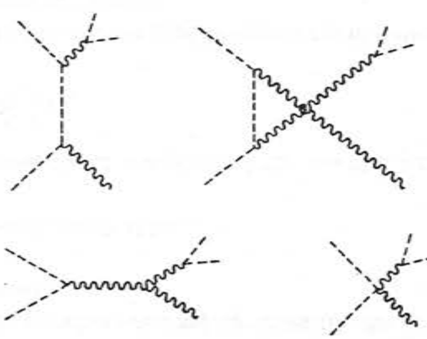

(6.58)

$j_{2\mu}(a_2)$ is not zero, so the process we have calculated does not seem to be gauge invariant. Is there a fault in the theory? Need the theory be gauge invariant if the $a_{2\mu}$ has come from a source? Yes, the theory must be – the fault lies in the fact that we are not now calculating a complete physical process. Consider as an example that the source of $a_{2\mu}$ is another scalon, i.e.



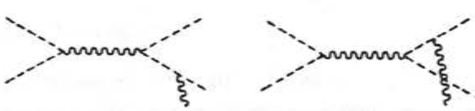

(6.59)

Then in addition to the diagrams



(6.60)

which contribute to (6.58), there are two additional diagrams which we must consider for the whole physical process, viz.



(6.61)

When the sum of all diagrams in (6.60) and (6.61) is considered and we put $a_{1\mu} = \alpha q_{1\mu}$, we will get zero. It is not surprising that we do not get zero for the diagrams (6.60) as they only constitute part of an answer to a physical process, and there is no reason why this partial answer should be gauge invariant. The diagrams (6.61) correspond to a modification of the source in



(6.58). Therefore, if we always only apply a gauge transformation to a physical process we will find that the amplitude is invariant (provided we stick with tree diagrams). Hence the diagram rules which we have constructed will work as long as we do not calculate any diagrams with closed loops. Before considering the difficulties with closed loops, we will consider the Compton scattering process further.

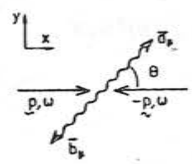
The formula for the real Compton effect is obtained from (6.51) by putting both gluons on their mass shells i.e.

$$q_1^2 = q_2^2 = 0; \qquad q_{1\mu}a_{1\mu} = q_{2\mu}a_{2\mu} = 0\,, \tag{6.62}$$

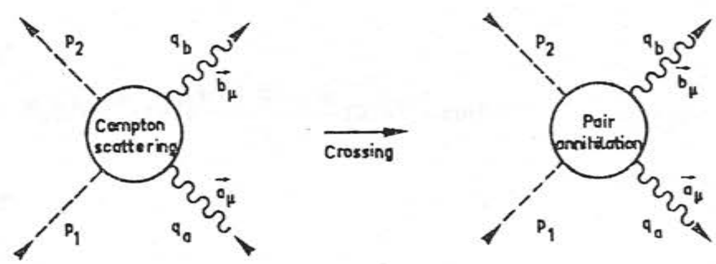
in accordance with the situation (a) earlier. We get

$$\frac{2g^2(S_2\times a_{2\nu})\cdot(S_1\times a_{1\mu})}{q_1\cdot q_2}\left\{-\frac{(p_{1\mu}p_{2\nu})(p_{1\lambda}q_{2\lambda})}{(p_1\cdot q_1)} + (p_{2\mu}p_{1\nu}) - \delta_{\mu\nu}(p_{1\lambda}q_{1\lambda})\right\} + \left\{\begin{matrix} S_1\leftrightarrow S_2\\ p_1\leftrightarrow -p_2\end{matrix}\right\}. \tag{6.63}$$

By taking the crossed diagrams of the Compton scattering process we obtain the process for scalon annihilation into gluons




in the center of mass system (6.64)




(6.65)

The amplitude for this process can be deduced immediately from (6.63); it is

$$\frac{2}{q_a\cdot q_b}\left[\frac{p_{1\mu}p_{2\nu}}{p_1\cdot q_a}+\frac{p_{1\nu}p_{2\mu}}{p_1\cdot q_b}+\delta_{\mu\nu}\right]\left[(p_1\cdot q_b)A_{\mu\nu}+(p_1\cdot q_a)B_{\mu\nu}\right], \tag{6.66}$$

where

$$A_{\mu\nu} = -(S_1 \times a_\mu)\cdot(S_2 \times b_\nu), \qquad B_{\mu\nu} = -(S_1 \times b_\nu)\cdot(S_2 \times a_\mu). \tag{6.67}$$

In the centre of mass system, the incident scalons have three momenta $p_i$ and $-p_i$ respectively and velocity $v_i$ and $-v_i$ where

$$v_i = p_i/\omega = p_i/\sqrt{p^2 + M^2}\,. \tag{6.68}$$

The amplitude for this process is then

$$\left[\frac{2v_i v_j}{1 - v^2\cos^2\theta} - \delta_{ij}\right]\left[(1 + v\cos\theta)A_{ij} + (1 - v\cos\theta)B_{ij}\right] \tag{6.69}$$

We consider the possible polarisation states for the gluon: state (a) in the plane of the reaction viz.,

$$a_\mu = \overset{t}{(0}, \overset{x}{-\sin\theta}, \overset{y}{\cos\theta}, \overset{z}{0)}\,\alpha\,, \tag{6.70}$$

state (b) perpensicular to the plane of the reaction viz.,

$$a_\mu = \overset{t}{(0}, \overset{x}{0}, \overset{y}{0}, \overset{z}{1)}\,\alpha\,. \tag{6.71}$$

Apply (6.69) to the following polarisation combinations:

(i) both gluons in state (b)

$$\text{Amplitude} = [(1 + v\cos\theta)A_{ii} + (1 - v\cos\theta)B_{ii}] \equiv \beta\,, \tag{6.72}$$

(ii) both gluons in state (a)

$$\text{Amplitude} = \beta t(\cos\theta)\,, \tag{6.73}$$

where

$$t(\cos\theta) = 1 - \frac{2\sin^2\theta\, v^2}{1 - v^2\cos^2\theta}\,, \tag{6.74}$$

(iii) one gluon in state (a) and one in state (b)

$$\text{Amplitude} = 0\,. \tag{6.75}$$

We can calculate $\beta$ given the isospin states of the scalons. The results are tabulated below, where $I$ is the total isospin of the scalons in the initial state

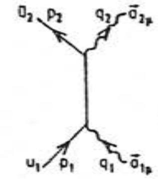
| $I$ | 0 | 1 | 2 |
|---|---|---|---|
| $\beta$ | 4 | $2v\cos\theta$ | 2 |

(6.76)



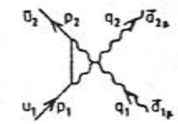
Finally we outline the calculation of the Compton scattering of gluons from isospin $\frac{1}{2}$ spinors using the Lagrangian (6.1) and the rules which we derived from it. The diagrams and the amplitudes are

(i)



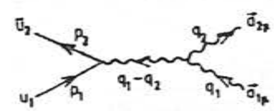
$$g^2\bar{u}_2(a_{2\mu}\cdot\tau\gamma_\mu)\frac{1}{\not p_1 + \not q_1 - m}(\gamma_\nu a_{1\nu}\cdot\tau)u_1\,, \tag{6.77}$$

(ii)



$$g^2\bar{u}_2(a_{1\nu}\cdot\tau\gamma_\nu)\frac{1}{\not p_1 - \not q_2 - m}(\gamma_\mu a_{2\mu}\cdot\tau)u_1 \tag{6.78}$$

(iii)

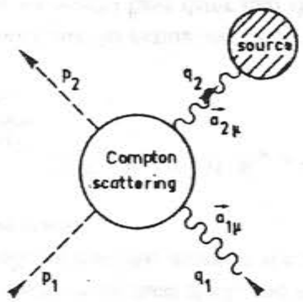


$$\frac{g^2\bar{u}_2\gamma_\mu\tau\cdot u_1}{(q_1 - q_2)^2}[(a_{2\nu}\times a_{1\mu})(2q_1 - q_2)_\nu + (a_{2\mu}\times a_{1\nu})(2q_2 - q_1)_\nu - (a_{2\nu}\times a_{1\nu})(q_1 + q_2)_\mu]\,. \tag{6.79}$$

We do not have the fourth diagram in this case. As in the scalar case, we can gauge one vector boson, $a_{1\mu} \to \alpha q_{1\mu}$, where upon we get, adding the results (6.77)–(6.79)

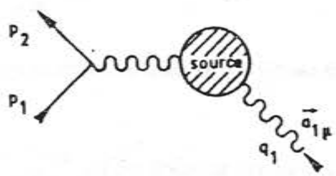
$$\frac{g^2\bar{u}_2\tau\cdot\gamma_\mu u_1}{(q_1 - q_2)^2}[j_{2\mu}(a_2)\times\alpha] \tag{6.80}$$

As before (6.57), $j_{2\mu}(a_2) = q_2^2 a_{2\mu} - q_{2\mu}(q_{2\nu}a_{2\nu})$ and again this is zero if $a_{2\mu}$ is free wave, or if it is a pure gradient (c.f. (6.54). However $j_{2\mu}(a_2)$ is not zero if $a_{2\mu}$ has a source ($j_{2\mu}(a_2) = \text{source}_{2\mu}$), but in this case, viz.,




(6.81)

We have again another type of diagram where the $a_{1\mu}$ acts on the source:



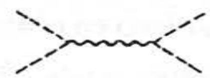

(6.82)

All the diagrams are taken into account, we again find that the physical process is gauge invariant. Hence there is no difficulty here as we expect, since these are tree diagrams.

### 7. Quantisation continued – Loops

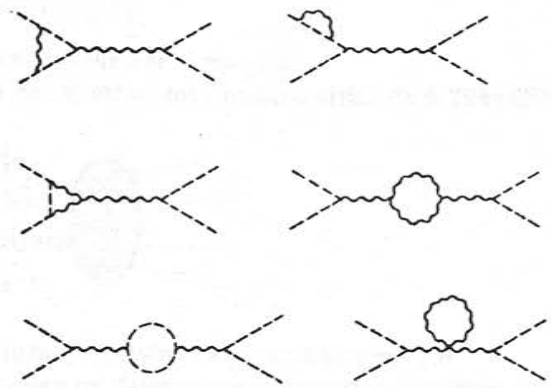
So far we have been successful in our attempt to quantise Yang–Mills theory. Now we want to try the procedure that we have adopted for higher order diagrams; this will entail diagrams with closed loops (the higher order tree diagrams do not present any difficulties). Why not therefore try to calculate the polarisation of the vacuum? But wait, if you want to discover a difficulty with a theory, you've got to look at a physical problem because some of your difficulties might come from not asking for a complete physical process (see the previous section). So the only way to discover whether something is right or wrong is not to pick up some arbitrary thing like the vacuum polarisation or the vacuum expectation value of a pair of operators, but to ask a physical question. But what should we ask? Correction to the Compton effect. This would be a fine problem, as good as any, but I decided to look at the following example – the scattering of two scalons by gluon exchange. In lowest order this is





(7.1)

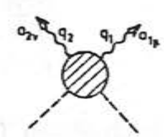
The correction diagrams to this coming from next order are of the type



(7.2)

We are not actually going to evaluate these diagrams, but only indicate what happens. When we calculate the diagrams (7.2) we, of course, have problems with renormalisation; but when we've straightened these out we would still like to check the result in some way. We can do this by checking whether or not we satisfy unitarity.

Consider the following diagram, where a pair of scalons annihilate into two gluons



(7.3)

The amplitude for this process is

$$\mathcal{M}_{\mu\nu} a_{1\mu} a_{2\nu} \, , \qquad (7.4)$$

where $\mathcal{M}_{\mu\nu}$ is some tensor function of the momenta and isospin labels of the scalons. (We calculated this process to first order in the previous chapter.)

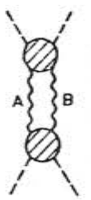
If we assume that the gluons are physical particles, there are constraints which their momenta and polarisation vectors must satisfy.

$$q_1^2 = q_2^2 = 0, \qquad q_1 \cdot a_1 = q_2 \cdot a_2 = 0\,, \tag{7.5}$$

i.e. there are only two polarisation states for each physical gluon. The probability for process (7.3) to take place is $\propto |\text{Amplitude}|^2$ i.e.

$$\text{Probability} \propto (\mathcal{M}_{\mu\nu} a_{1\mu} a_{2\nu})(\mathcal{M}^\dagger_{\mu'\nu'} a_{1\mu'} a_{2\nu'}) \tag{7.6}$$

where we have taken the $a$'s to be real. In calculating this we see that it is closely related to the amplitude for the following process




(7.7)

The amplitude for this process is

$$\sum_{\substack{\text{all } \mu \\ \text{all } \nu}} \mathcal{M}_{\mu\nu} \mathcal{M}^\dagger_{\mu\nu} \times (\text{propagator terms for gluons A and B})\,. \tag{7.8}$$

If we take the imaginary part of (7.8) we get

$$\sum_{\substack{\text{all } \mu \\ \text{all } \nu}} \mathcal{M}_{\mu\nu} \mathcal{M}^\dagger_{\mu\nu}\,, \tag{7.9}$$

where the propagators have been removed by this process and the corresponding gluons are now on mass shell, and irrelevant constants have been ignored. Unitarity tells us that this must be the same as (7.6) summed over physical polarisation states:

$$\sum_{\substack{\text{physical} \\ \text{polarisation} \\ \text{states}}} (\mathcal{M}_{\mu\nu} a_{1\mu} a_{2\nu})(\mathcal{M}_{\mu'\nu'} a_{1\mu'} a_{2\nu'})\,. \tag{7.10}$$

However note that in evaluating (7.9) we have summed over more polarisation states than we would have done had the intermediate gluons been physically real particles, so it is not obvious that (7.9) and (7.10) are equal.

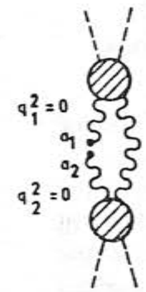
In making the comparison we will only consider the $\mu$ index. We lose no



generality by doing this since if we can sort this out then we can also sort out the $\nu$ index. In making this assumption, we really compare

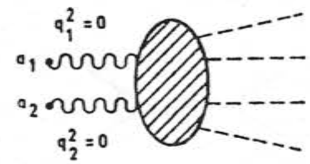
$$\sum_{\substack{\text{2 poln. states} \\ \text{of } a_{1\mu};\text{ all} \\ \nu \text{ states}}} \mathcal{M}_{\mu\nu} a_{1\mu} \mathcal{M}^\dagger_{\mu'\nu} a_{1\mu'}\,, \tag{7.11}$$

with (7.9). We can think of (7.11) as follows




(7.12)

The gluons $a_1$ and $a_2$ are on mass shell since they are physical gluons. This diagram, when the gluon line is not connected, is similar to, but more complicated than the Compton scattering process; it is




(7.13)

When we calculated the Compton effect we noticed that when $a_{1\mu} = \alpha q_{1\mu}$, then the amplitude was of the form

$$g(\alpha \times c_\lambda) \cdot j_\lambda(a_2)\,, \tag{7.14}$$

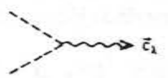
where

$$j_\lambda(a_2) = q_2^2 a_{2\lambda} - (q_{2\nu} a_{2\nu}) q_{2\lambda}\,. \tag{7.15}$$

In the case of the Compton scattering proper

$$c_\lambda = \frac{(S_1 \times S_2)(p_1 + p_2)_\lambda}{(p_1 - p_2)^2} \tag{7.16}$$

but in general we may think of $c_\lambda$ as being the gluon field produced by some source


(7.17)

Since we are really only interested in the indices, define

$$\widetilde{\mathcal{M}}_{\mu\mu'} = \sum_{\text{all } \nu} \mathcal{M}_{\mu\nu} \mathcal{M}^{\dagger}_{\mu'\nu} , \tag{7.18}$$

so that (7.11) becomes

$$\sum_{\substack{\text{2 poln. states}\\ \text{of } a_{1\mu}}} \widetilde{\mathcal{M}}_{\mu\mu'} a_{1\mu} a_{1\mu'} , \tag{7.19}$$

and (7.9) becomes

$$\sum_{\text{all } \mu} \widetilde{\mathcal{M}}_{\mu\mu} \tag{7.20}$$

Since $q_1^2 = 0$ we can choose the axes so that

$$q_1 = \omega(\overset{t}{1}, \overset{z}{1}, \overset{x}{0}, \overset{y}{0}) . \tag{7.21}$$

$a_1$ then has components only in the $x$ and $y$ direction and (7.19) becomes

$$-\widetilde{\mathcal{M}}_{xx} - \widetilde{\mathcal{M}}_{yy} . \tag{7.22}$$

Similarly (7.20) becomes

$$\widetilde{\mathcal{M}}_{tt} - \widetilde{\mathcal{M}}_{xx} - \widetilde{\mathcal{M}}_{yy} - \widetilde{\mathcal{M}}_{zz} . \tag{7.23}$$

Eqs. (7.22) and (7.23) differ by

$$\widetilde{\mathcal{M}}_{zz} - \widetilde{\mathcal{M}}_{tt} . \tag{7.24}$$

So when we took the imaginary part of (7.8) we obtained the extra piece (7.24), which we do not want; the theory will be unitary if we can show that this extra piece is zero. Rotating the axes we can re-write (7.24) as

$$\tfrac{1}{2}\{\widetilde{\mathcal{M}}_{(z-t),(z+t)} + \widetilde{\mathcal{M}}_{(z+t),(z-t)}\} . \tag{7.25}$$

Consider the first term in (7.25): the second term is similar. We would get such a term if we considered a process with two gluons which have polarisations in the directions



$$\frac{1}{\sqrt{2}}(\overset{t}{1}, \overset{z}{1}, \overset{x}{0}, \overset{y}{0}) \quad \text{viz.} \quad a_{1\mu} = q_{1\mu} , \tag{7.26}$$

and

$$\frac{1}{\sqrt{2}}(1, -1, 0, 0) \quad \text{viz.} \quad a_{2\mu} \equiv \frac{N_\mu}{q_2 \cdot N} , \tag{7.27}$$

We get zero for $\widetilde{\mathcal{M}}_{(z-t),(z+t)}$ only if $j_\lambda(a_2)$ is zero (from (7.14)). However it is not zero although $q_2^2$ is, since $q_2 \cdot a_2$ is not zero (the particle is not free).

In QED, this problem does not occur, because when we have a process with two external photons, if we gauge one photon (i.e. put the polarisation vector parallel to the momentum) the amplitude vanishes irrespective of whether or not the other photon is physical. That is, the quantity analogous to $j_\lambda(a_2)$ is identically zero independently of the state of the photon.

There is nothing we can do about this problem in Yang–Mills theory. If we wish to make the theory unitary we must subtract something to get rid of these extra pieces. By taking the simple Compton case, we can get some idea of the form of the thing we have to subtract. The Compton amplitude (7.14) becomes

$$g(\boldsymbol{\alpha} \times c_\lambda) \cdot \boldsymbol{\alpha} q_{2\lambda} , \tag{7.28}$$

when we use (7.27) and the fact that $q_2^2 = 0$. The polarisation $\boldsymbol{\alpha}$ is summed over when we make a closed loop. This is the extra piece that we have to get rid of. We can see that the most direct way to do this is to add an isovector scalar particle $\boldsymbol{\alpha}$ which is self coupled and also coupled to the vector according to (7.28). (This is the ghost particle.) In co-ordinate space this coupling becomes

$$g(P \times A_\lambda) \cdot \partial_\lambda P . \tag{7.29}$$

Clearly since the particle has to cancel a piece coming from the $A_\mu$, its propagator must have the same form as that of the $A_\mu$ viz. $1/k^2$. We can see from the form of its coupling that this ghost appears in closed loops in diagrams in the same way as $A$'s. It is not allowed to appear in the initial or final states since it is not a physical particle. It was not obvious to me from the type of analysis presented above, what to do in a diagram with more than one closed loop. (Unitarity is not a sufficient constraint in this case.) In fact the solution to the problem for larger numbers of loops is to add the ghost as if it were a Fermi particle. (At the one loop level, adding a Fermi particle is equivalent to subtracting the contribution from a particle coupled in the same way but having Bose statistics, which is what we did above.) The contribution from this ghost could come from a Lagrangian of the form

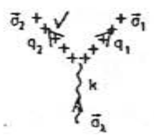
$$\mathcal{L}_g = \tfrac{1}{2}\partial_\mu P^+ \cdot \partial_\mu P + g\partial_\mu P^+ \cdot (A_\mu \times P) \tag{7.30}$$

This Lagrangian leads to the following diagram rules.

ghost propagator:

$$\text{a xxxxxxxxx b} \qquad \delta_{ab}/k^2 \tag{7.31}$$

ghost vertex:



$$ig q_{2\lambda}\sigma_2 \cdot (a_\lambda \times \sigma_1)\,, \tag{7.32}$$

where the ghost only enters in closed loops. Topologically for every diagram with an $A_\mu$ closed loop there is one with a ghost loop in the same place. Note that (7.32) is not symmetric looking and this asymmetry is fundamental, so we add a check mark (√) to one of the ghost legs. We must keep track of these √'s as we go round a closed loop so as to ensure that they are always on the same side of a vertex. (It does not matter on which leg we put the √ provided that we always put it on the same one, since it can be shown that both choices lead to the same results.) In order to get the right factors it is necessary to assume that the ghost is a complex field.

All these details were proved by Fadde'ev and Popov [26]. We haven't proved them; we have merely indicated how if we proceed in a straight forward and naive way, we come into difficulties; and we got a pretty good smell of what to do. But for the full glory of the theory we must follow the approach of Fadde'ev – it requires more machinery and we will discuss it later.

Finally we write the total Lagrangian as

$$\mathcal{L} = -\tfrac{1}{4}E_{\mu\nu}\cdot E_{\mu\nu} - \tfrac{1}{2}(\partial_\mu A_\mu)^2 + \tfrac{1}{2}\partial_\mu P^+ \cdot \partial_\mu P + \tfrac{1}{2}g\partial_\mu P^+ \cdot (A_\mu \times P) + \text{Matter}(\phi, \psi, A) \tag{7.33}$$

Why is the second term in (7.33) present? If we expand the quadratic parts of $-\tfrac{1}{4}E_{\mu\nu}\cdot E_{\mu\nu}$ which will give us the propagator we get:

$$-\tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)\cdot(\partial_\mu A_\nu - \partial_\nu A_\mu) = -\tfrac{1}{2}(\partial_\mu A_\nu)^2 + \tfrac{1}{2}(\partial_\mu A_\mu)^2\,. \tag{7.34}$$

We have changed the form of the Lagrangian in this by integrating the action by parts, throwing away (as usual) the surface terms at infinity. These changes



can produce no physics as they do not change the action. We now see that the $-\tfrac{1}{2}(\partial_\mu A_\mu)^2$ term in (7.33) cancels the same term in (7.34) to leave the quadratic terms in the Lagrangian as

$$-\tfrac{1}{2}(\partial_\mu A_\nu)^2\,. \tag{7.35}$$

This is exactly what we want to give the propagator $\delta_{ab}\delta_{\mu\nu}/k^2$ which we have been using. (This is equivalent to choosing the gauge $\partial_\mu A_\mu = 0$.) That the Lagrangian (7.33) is equivalent to the Lagrangian

$$\mathcal{L} = -\tfrac{1}{4}E_{\mu\nu}\cdot E_{\mu\nu} + \text{Matter}(\phi, \psi, A) \tag{7.36}$$

(i.e. that the ghost Lagrangian (7.30) compensates for the extra term) is a more difficult problem which was solved by Fadde'ev and Popov [26], who also showed what form the extra terms in the Lagrangian would take in different gauges.

## 8. Quantisation via the Hamiltonian formalism. A ghostless gauge

We will now quantise Yang–Mills theory by going via the Hamiltonian formalism, rather than by path integrals; a similar procedure has been carried out in QED and one disadvantage is that we lose manifest Lorentz invariance, although we know that the final results must be relativistically invariant. We must choose a gauge before constructing the Hamiltonian since, as we know from our experience with QED, otherwise there are too many variables. The Hamiltonian is a machine for telling us how operators change with time; thus the important operator is

$$\partial/\partial t\,. \tag{8.1}$$

But we have seen that in Yang–Mills theory the physically interesting operator is $D_\mu$ which has a time component

$$\left(\frac{\partial}{\partial t} + A_0 \times\right). \tag{8.2}$$

Hence to recover the operator (8.1) which is appropriate for the Hamiltonian formalism, we want to put

$$A_0 = 0\,. \tag{8.3}$$

Since we have the freedom of gauge transformations we can make this choice. If we do not do this the Hamiltonian formalism becomes extremely cumbersome and very difficult. By the way, why don't we make this choice in QED?

If we do, then a static charge has a vector potential associated with it which rises linearly with time. This is an inconvenience. However, in QCD, perhaps there are no static charges, i.e. we imagine that hadrons cannot separate into quarks or, when doing a problem, we start initially with no matter present and then create the matter and anti-matter (so that by doing this we ensure that the total colour charge of the world stays zero) which exists over a finite time. There seems to be a danger in this non-linear theory, since if the vector potential rises linearly with time, certain cross terms may produce physical effects (e.g. by quantum fluctuations); of course that's a dream, since if it were true, we could say that to avoid these difficulties, the vector potential is not allowed to rise linearly with time for ever, and hence we have proved colour confinement.

We shall write the classical equations for the Yang–Mills field, construct the Hamiltonian, having defined the conjugate momenta, and then quantise the theory by using the canonical commutation relations. The classical equation of motion is (2.28)

$$\partial_\mu E_{\mu\nu} = -J_\nu^{\text{matter}} - A_\mu \times E_{\mu\nu} ,$$

where

$$E_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + A_\mu \times A_\nu , \tag{8.4}$$

The $E_{\mu\nu}$ tensor has six (Lorentz) components. Since we have singled out the time axis by our gauge choice, consider the $ti$ component of $E_{\mu\nu}$. ($i = x, y, z$). From (8.4)

$$E_{ti} = \frac{\partial A_i}{\partial t} \equiv \mathcal{E}_i , \tag{8.5}$$

where by definition $\mathcal{E}_i$ is the colour electric field (named using the analogy with QED). Similarly the $ij$ components are (from (8.4))

$$E_{ij} = \partial_i A_j - \partial_j A_i + A_i \times A_j \equiv \epsilon_{ijk} B_k , \tag{8.6}$$

where by definition $B_k$ is the colour magnetic field. Using this definition, (8.6) gives

$$\underset{\sim}{B} = \text{curl}\, \underset{\sim}{A} + \tfrac{1}{2} \underset{\sim}{A} \overset{\times}{\times} \underset{\sim}{A} . \tag{8.7}$$

Look at the $i$ component of (8.4). This now reads

$$\frac{\partial^2 A_i}{\partial t^2} + (\text{curl}\, \underset{\sim}{B})_i = -J_i - (\underset{\sim}{A} \overset{\times}{\times} \underset{\sim}{B})_i , \tag{8.8}$$

where we have introduced the notation that the upper vector symbols refer to isospace and the lower symbols to real space.



These three equations can be written generally as

$$\frac{\partial^2 \underset{\sim}{A}}{\partial t^2} + \text{curl}\, \underset{\sim}{B} = -\underset{\sim}{J} - \underset{\sim}{A} \overset{\times}{\times} \underset{\sim}{B} . \tag{8.9}$$

The fourth equation comes from taking the time component of (8.3); this is

$$\frac{\partial}{\partial t}(\underset{\sim}{\nabla} \cdot \underset{\sim}{A}) = -\rho - \underset{\sim}{A} \overset{\times}{\cdot} \underset{\sim}{\mathcal{E}} , \tag{8.10}$$

where we have put

$$\rho \equiv J_0 . \tag{8.11}$$

If we solve (8.9) we do not have to solve (8.10), because (8.10) is practically a consequence of (8.9); we can see this by taking the divergence of (8.9) and re-arranging the terms:

$$\frac{\partial}{\partial t}\left[\frac{\partial}{\partial t} \underset{\sim}{\nabla} \cdot \underset{\sim}{A}\right] = \frac{\partial}{\partial t}\left[-\underset{\sim}{A} \overset{\times}{\cdot} \underset{\sim}{\mathcal{E}}\right] - \underset{\sim}{\nabla} \cdot \underset{\sim}{j} - \underset{\sim}{A} \overset{\times}{\cdot} \underset{\sim}{j} . \tag{8.12}$$

This is the same as the time derivative of 8.10 provided

$$\frac{\partial \rho}{\partial t} = \underset{\sim}{\nabla} \cdot \underset{\sim}{j} + \underset{\sim}{A} \overset{\times}{\cdot} \underset{\sim}{j} . \tag{8.13}$$

The matter will be such that we have current conservation (in the covariant sense) and hence this equation will be satisfied. In principle, integrating (8.12) with respect to time to get (8.10) (using (8.13)), we could generate a function of $x, y$, and $z$. We assume that the initial conditions are such that this function is zero; it will then stay zero for all time. We can therefore forget about (8.10) in solving the theory.

We can solve (8.9) for $A_i(\underset{\sim}{x}, t)$: (8.9) is just a complicated non-linear differential equation for $A_i$. The dynamical features of the theory are simple since the only derivative of $A_i$ which appears is the second time one, and this appears linearly. Can we find a Lagrangian from which (8.9) comes? It's easy, because we know all about Newton's laws and to get a term like $\partial^2 A_i/\partial t^2$ in the equation of motion, we know that we must have a term like $(\partial A_i/\partial t)^2$ in the Lagrangian. This is just like having the square of the velocity in the Lagrangian; the equation of motion then has a term linear in the acceleration. In order to generate the rest of the terms in the equation of motion we proceed as follows: the term $J_i$ is generated by a Lagrangian term of the form $J_i \cdot A_i$, and it turns out that the remaining field pieces i.e. $\underset{\sim}{\nabla} \times \underset{\sim}{B} + \underset{\sim}{A} \overset{\times}{\times} \underset{\sim}{B}$ can be derived from a term of the form $\underset{\sim}{B} \overset{\cdot}{\cdot} \underset{\sim}{B}$. So the Lagrangian has the form

$$-\frac{1}{2}\frac{\partial A_i}{\partial t} \cdot \frac{\partial A_i}{\partial t} - \tfrac{1}{2}B_i \cdot B_i + J_i \cdot A_i \,. \tag{8.14}$$

This is clearly the same as the Lagrangian

$$-\tfrac{1}{4}E_{\mu\nu} \cdot E_{\mu\nu} + \text{Matter}\,, \tag{8.15}$$

when we put $A_0 = 0$ and use the definition of $B_i$ given earlier. What about the Hamiltonian? Nothing could be simpler; define the momentum conjugate to $A_i$ by

$$\Pi_i = \frac{\delta \mathcal{L}}{\delta \dot{A}_i} = -\frac{\partial A_i}{\partial t}. \tag{8.16}$$

To get the Hamiltonian take the square of the momentum and add the potential energy, viz.

$$H = \tfrac{1}{2}\Pi_i \cdot \Pi_i + \tfrac{1}{2}B_i \cdot B_i - J_i \cdot A_i \,. \tag{8.17}$$

We now quantise the theory by imposing the canonical quantisation conditions, and treating $\Pi$ and $A$ as operators

$$[\Pi_i^a(\underset{\sim}{x}, t), A_j^b(\underset{\sim}{y}, t)] = i\delta^{ab}\delta^{ij}\delta^3(\underset{\sim}{x} - \underset{\sim}{y})\,. \tag{8.18}$$

What would the diagrams look like for such a theory? We can get the propagator from (8.9) by looking at the linear part. This is substituting for $\underset{\sim}{B}$ using (8.7)

$$\frac{\partial^2 \underset{\sim}{A}}{\partial t^2} + \text{curl}\,(\text{curl}\,\underset{\sim}{A}) = -\underset{\sim}{J} - \underset{\sim}{A} \overset{\times}{\times} \underset{\sim}{B} - \tfrac{1}{2}\,\text{curl}\,(\underset{\sim}{A} \overset{\times}{\times} \underset{\sim}{A})\,. \tag{8.19}$$

All the terms on the left side are linear in $\underset{\sim}{A}$ and all the terms on the right are higher order. We can solve this to obtain the propagator which is

$$\frac{\delta_{ij} - (k_i k_j/\omega^2)}{\omega^2 - k_i k_i}\,. \tag{8.20}$$

This proof is left as an exercise; it is simple to complete using the method we described in sect. 6.

What about the coupling terms? They are the same as before except that there are no time components. This set of rules then gives the correct answers for Yang–Mills problems.

No fooling around! No ghosts! No nothing! Isn't that wonderful? Yes it's wonderful, but it's also mysterious because all the rules depend on choosing a time axis and it's not clear that the answer to a physical problem will be relatavistically invariant. So what? We know that it must be relativistically invariant because the theory we started with was. True, but the theory diverges and we encounter the delicate problem of imposing renormalisation without destroying Lorentz invariance – a problem which took 20 years to solve in the case of QED. It shouldn't take us this long to solve the problem in Yang–Mills as we're much smarter now.

Despite these problems these are perfectly good rules and we can go away and calculate with them.

Suppose we try to compare closed loops with this propagator and with the propagator we had earlier viz.

$$\delta_{\mu\nu}/k^2\,. \tag{8.21}$$

In using (8.21) we are counting the contribution from $A_0$; but there are superfluous degrees of freedom; therefore we must subtract something to get rid of them – this something is the ghost. In some sense therefore, the ghost is the difference between the result one gets by using (8.20) and (8.21). What we subtract doesn't look relativistically invariant. The subtraction will be different depending on the orientation of the axes; in this case we have singled out the time axis and therefore if we rotate the axes things will change. The problem of showing that the results are independent of the method of subtraction was solved by Fadde'ev and Popov [26].

## 9. The equivalence of different gauges

The purpose of this chapter is to make clear the ideas involved in Fadde'ev and Popov's demonstration of the equivalence of different gauges. I shall not prove anything in great detail as this would take too long. We will start with the formalism of sect. 8, the $A_0 = 0$ gauge, and attempt to transform it so as to obtain the formalism in different gauges; in particular we will try to obtain the system in which the propagator is $\delta_{\mu\nu}/k^2$ ($\partial_\mu A_\mu = 0$ gauge). The simplest thing to do would be to write down the rules in the $A_0 = 0$ gauge and then transform them to the rules for the $\partial_\mu A_\mu = 0$ gauge. In order to get more generality and check that the formalism works in all orders we will use a more central approach via path integrals. What we shall do is to construct a path integral for the $A_0 = 0$ gauge, transform the path integral to the other gauge, and then get the rules for the other gauge from the path integral. We shall see that these rules are the same as those obtained in sects. 6 and 7.

As everybody knows, if we have a Lagrangian of the form

$$L = \frac{m\dot{q}^2}{2} - V(q) \tag{9.1}$$

Then the amplitude for the particle to go along a particular trajectory $q(t)$ is [18]

$$\exp(\mathrm{i}S/\hbar)\,. \tag{9.2}$$

where $S$ is the classical action obtained from (9.1). The total amplitude for the particle to go from $a$ to $b$ is obtained by summing over all trajectories connecting $a$ and $b$. viz.

$$\text{Amplitude} \propto \int \exp\left\{\frac{\mathrm{i}}{\hbar}\int \left[\frac{m\dot{q}(t)^2}{2} - V(q(t))\right] \mathrm{d}t\right\} \mathcal{D}q(t)\,. \tag{9.3}$$

If we have a system of several variables $q_i(t)$, (9.3) becomes

$$\text{Amplitude} \propto \int \exp\left\{\frac{\mathrm{i}}{\hbar}\int \sum_i \left[\frac{m_i\dot{q}_i(t)^2}{2} - V(q_i(t))\right] \mathrm{d}t\right\} \mathcal{D}q_i(t)\,. \tag{9.4}$$

In a field theory the variable $i$ becomes continuous and is denoted by $\underset{\sim}{x}$. Eq. (9.4) then becomes

$$\text{Amplitude} \propto \int \exp\left\{\frac{\mathrm{i}}{\hbar}\int \mathcal{L}(A)\,\mathrm{d}^3\underset{\sim}{x}\,\mathrm{d}t\right\} \mathcal{D}A(\underset{\sim}{x}, t)\,, \tag{9.5}$$

where $\mathcal{L}(A)$ is the Lagrangian density for the theory and the integral over $\underset{\sim}{x}$ replaces the sum over $i$ in the limit $i$ continuous

$$\mathcal{L}(A) = \left(\frac{\partial A(\underset{\sim}{x}, t)}{\partial t}\right)^2 - V(A)\,. \tag{9.6}$$

Eq. (9.5) means this: if $A$ is distributed in some way in spacetime, there is a certain action for this distribution which we can calculate. The classical theory corresponds to the distribution which makes the action a minimum. Quantum mechanics corresponds to adding the amplitude corresponding to the actions for all possible distributions. We can deduce the classical limit by noticing that in the integration over field distributions, the amplitudes corresponding to near stationary values of the action add coherently whereas those far from the minimum are changing so rapidly that the amplitudes from neighbouring distributions tend to cancel. In the limit of action $S/\hbar$ being very large only paths near the minimum contribute to the integral over distributions and so we obtain the classical theory.

We now return to the Lagrangian density which we derived for the $A_0 = 0$ gauge in the previous section (8.14) and substitute it in (9.5). This gives

$$\int \exp\left\{\mathrm{i}\int \left[-\frac{1}{2}\frac{\partial \underset{\sim}{A}(\underset{\sim}{x}, t)}{\partial t} : \frac{\partial \underset{\sim}{A}(\underset{\sim}{x}, t)}{\partial t} - \tfrac{1}{2}\underset{\sim}{B} : \underset{\sim}{B} + \underset{\sim}{J} : \underset{\sim}{A}\right] \mathrm{d}^4x\, \mathcal{D}\underset{\sim}{A}(\underset{\sim}{x}, t)\,\mathcal{D}_{\text{Matter}}\,. \tag{9.7}$$



$\mathcal{D}\underset{\sim}{A}(\underset{\sim}{x}, t)$ is defined to mean $\Pi_{i,a,x,t}\,\mathrm{d}A_i^a(\underset{\sim}{x}, t)$ and in $\mathcal{D}_{\text{Matter}}$ we have to integrate over the spin $\frac{1}{2}$ $\psi$ field. This means that we must use the path integral using Grassman algebra, which we mentioned earlier.

Unfortunately from the path integral viewpoint, the expression is not manifestly Lorentz invariant. However we can write (9.7) as follows

$$\int \exp\left\{\mathrm{i}\int \left[-\tfrac{1}{4}E_{\mu\nu}\cdot E_{\mu\nu} + J_\mu \cdot A_\mu\right]\mathrm{d}^4x\right\} \delta(A_0(\underset{\sim}{x}, t))\,\mathcal{D}A_\mu(x, t)\,\mathcal{D}_{\text{Matter}}\,, \tag{9.8}$$

where $\delta(A_0(\underset{\sim}{x}, t))$ is defined to mean $\Pi_{a,x,t}\,\delta(A_0^a(\underset{\sim}{x}, t))$ and in this expression the functional integral runs over all four space-time components of $A_\mu(\underset{\sim}{x}, t)$. Although (9.8) and (9.7) are equivalent (by construction) we note that in (9.8):

(i) $E_{\mu\nu}\cdot E_{\mu\nu}$ is invariant under all gauge transformations;

(ii) $J_\mu \cdot A_\mu$ is also invariant under gauge transformations provided that the matter fields transform in the correct way;

(iii) $\mathcal{D}A_\mu$ is also invariant under an infinitesimal gauge transformation

$$A_\mu \to A_\mu + \partial_\mu \boldsymbol{\alpha} + \boldsymbol{\alpha} \times A_\mu\,. \tag{9.9}$$

We can see that (iii) is true by the following argument. Since we are considering $\mathcal{D}A_\mu$, i.e. an infinitesimal, we can drop the $\partial_\mu\boldsymbol{\alpha}$ term since it is higher order in infinitesimal. $\mathcal{D}A_\mu$ is a volume element in isospace

$$\mathcal{D}A_\mu = \prod_{\mu, \underset{\sim}{x}, t} \mathrm{d}A_\mu^1(\underset{\sim}{x}, t)\,\mathrm{d}A_\mu^2(\underset{\sim}{x}, t)\,\mathrm{d}A_\mu^3(\underset{\sim}{x}, t)\,. \tag{9.10}$$

The remaining gauge transformation in (9.9) ($\boldsymbol{\alpha} \times A_\mu$) is a rotation in isospace and volume elements are invariant under such transformations (cf. $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ which is invariant under a rotation in three dimensional Euclidean space). Hence $\mathcal{D}A_\mu$ is invariant.

We have therefore shown that (9.8) is gauge invariant apart from the $\delta$-function. What happens to it under a gauge transformation? Under a gauge transformation (not necessarily infinitesimal),

$$A_0 \to GA_0\,, \tag{9.11}$$

which in general contains the space as well as time components of $A_\mu$. Since any $A_\mu$ distribution can be gauge transformed so that $A_0 = 0$, the $\delta$-function is really more universal than it looks. Consequently we can get to any other gauge by suitably choosing $G$.

As a specific example, let us choose to transform to the gauge

$$\partial_\mu A_\mu = 0 . \tag{9.12}$$

We will find that when we transform to this gauge,

$$\delta(A_0) \to \delta(\partial_\mu A_\mu)\mathcal{F}(A) \tag{9.13}$$

and we will also find that the factor $\mathcal{F}(A)$ can be written in such a way that it exactly reproduces the ghost term (Lagrangian (7.30)) which we required in the $\partial_\mu A_\mu = 0$ gauge.

There are several mathematical tricks which we will need in what follows. Because of the vast number of indices flying around, we will often use a symbolic notation.

*Mathematical trick 1.* Since almost all the questions about path integrals which are important are concerned with changes in the fields, the overall normalisation of the path integrals is irrelevant. We can, therefore, ignore any constants (including infinite ones) which appear as multiplicative factors outside the path integral provided that they do not depend on the fields.

*Mathematical trick 2.* Suppose that we have a number of variables $x_i$ and a number of functions $y_i$, which depend on the $x_i$. If we wish to transform from the $x_i$ to the $y_i$ then the δ-function transform as follows:

$$\delta(y_i - y_i(0)) = \left[\operatorname{Det}\frac{\partial y_i}{\mathrm{d}x_j}\right]^{-1}_{x_j=0} \delta(x_j) , \tag{9.14}$$

or

$$\delta(x_j) = \left[\operatorname{Det}\frac{\partial y_i}{\mathrm{d}x_j}\right]_{x_j=0} \delta(y_i - y_i(0)) ,$$

where as before the δ-functions are the generalised functions, i.e. $\delta(y_i)$ means $\Pi_i \delta(y_i)$. This can be seen by extending the result for one variable $y = f(x)$ viz.

$$\delta(y) = \frac{1}{|f'(x)|}\delta(x) , \tag{9.15}$$

to several variables.

*Mathematical trick 3.* Consider the integral

$$\int \exp(-\tfrac{1}{2}cp^2)\,\mathrm{d}p = \sqrt{2\pi/c} . \tag{9.16}$$

We drop the $\sqrt{2\pi}$ factors in what follows since they will be irrelevant because of mathematical trick 1. Let us now have several variables and look at the integral



$$\int \exp\left(-\tfrac{1}{2} \sum_i p_i c'_i p_i\right) \prod_i \mathrm{d}p_i . \tag{9.17}$$

This clearly has the value $1/\sqrt{\Pi_i c'_i}$.

If we do a linear transformation on these $p_i$'s:

$$p_i \to \sum_j a_{ij} p_j , \tag{9.18}$$

then (9.17) becomes

$$\int \exp\left(-\tfrac{1}{2} \sum_{ij} p_i c_{ij} p_j\right) \prod_i \mathrm{d}p_i = \frac{1}{\sqrt{\operatorname{Det} c_{ij}}} \tag{9.19}$$

where the $c$'s and the $c'$'s are related by (9.18). Eq. (9.19) is like a path integral in quantum mechanics for a Bose particle interacting in a way described by the matrix $c_{ij}$ (provided that we make a Wick rotation of the time axis $\mathrm{d}t \to \mathrm{i}\mathrm{d}t$). If we suppose that $p_i$ is complex then we are really integrating over twice as many variables and the square root disappears:

$$\int \exp\left(-\tfrac{1}{2} \sum_{ij} p_i^* c_{ij} p_j\right) \prod_i \mathrm{d}p_i = \frac{1}{\operatorname{Det} c_{ij}} . \tag{9.20}$$

If we try to do the same integral for particles $Q_i$ with Fermi statistics (i.e. we introduce a Grassman algebra), we find that, with a suitable definition of what we mean by an integral in this case,

$$\int \exp\left\{-\tfrac{1}{2} \sum_{ij} Q_i^* c_{ij} Q_j\right\} \prod_i \mathrm{d}Q_i = \operatorname{Det} c_{ij} . \tag{9.21}$$

It will turn out that (9.21) is just the reason that the ghost has Fermi statistics.

When we transform from the $\delta(A_0)$ to the $\delta(\partial_\mu A_\mu)$ in the functional integral, we will pick up a Jacobian similar to that in (9.14). Symbolically we can think of the variable $x_i$ as being $A_0(\underset{\sim}{x}, t)$ and $y_i$ as being $\partial_\mu A_\mu(\underset{\sim}{x}, t)$ (in this case $y_i(0)$ is zero). We will then use (9.21) to turn the determinant into a path integral. Another equation is useful when it comes to finding the coefficients $c_{ij}$; consider

$$y_i(x_j + p_j) = y_i(x_j) + c_{ij} p_j , \tag{9.22}$$

where we have defined $c_{ij} = (\mathrm{d}y_i/\mathrm{d}x_j)_{x_j=0}$. On multiplying by $p_i^*$ we get

$$p_i^* [y_i(x_j + p_j) - y_i(x_j)] = p_i^* c_{ij} p_j . \tag{9.23}$$

The term on the right has the same form as the expression in (9.20). We are trying to transform from the gauge $A_0 = 0$ to the gauge $\partial_\mu A_\mu = 0$ and therefore we need the matrix

$$\Delta(\partial_\mu A_\mu)/\Delta(A_0)\,. \tag{9.24}$$

which is analogous to the matrix $dy_i/dx_j$. This matrix $\Delta(\partial_\mu A_\mu(y))/\Delta(A_0(x))$ has infinite dimension corresponding to the continuous variables $x$ and $y$, and in addition has isospin labels. If we try to evaluate (9.24) directly, we will run into difficulties since the two gauges are linked by a finite gauge transformation which depends on the original fields. (We know that finite gauge transformations are difficult to handle.) The derivative $dy_i/dx_j$ can fortunately be obtained in an indirect way. Suppose $y_i$ and $x_j$ are changed by changing a variable $t$. We may be able to evaluate $dy_i/dx_j$ by first evaluating $dy_i/dt$ and $dx_j/dt$ and then using

$$\frac{dy_i}{dx_j} = \frac{dy_i}{dt} \cdot \frac{dt}{dx_j}. \tag{9.25}$$

It was the very clever idea of Fadde'ev to extend this idea to evaluate (9.24): in this case he took $t$ to be an infinitesimal gauge transformation. If we do an infinitesimal gauge transformation on $A_0$ (gauge parameter $\boldsymbol{\alpha}$) we get:

$$A_0' = A_0 + \boldsymbol{\alpha} \times A_0 + \partial_0 \boldsymbol{\alpha}\,. \tag{9.26}$$

Transform $\partial_\mu A_\mu$ by an infinitesimal gauge transformation with parameter $\boldsymbol{\beta}$:

$$\partial_\mu A_\mu' = \partial_\mu A_\mu + \boldsymbol{\beta} \times \partial_\mu A_\mu + \partial_\mu \beta \times A_\mu + \partial_\mu \partial_\mu \boldsymbol{\beta}\,. \tag{9.27}$$

As we are trying to use an equation of the form (9.25) to evaluate the matrix (9.24), we would naively think that we should take $\boldsymbol{\alpha} = \boldsymbol{\beta}$. But wait a minute. What we really need is for the gauge parameter $\boldsymbol{\beta}$ to be $\boldsymbol{\alpha}$ gauge transformed by the finite gauge transformation $G$, defined by (9.11). (All the fields are rotated by the gauge transformation $G$ and so what we really want is to know how the fields $A_0$ and $\partial_\mu A_\mu$ change in terms of the variables $\boldsymbol{\alpha}$ and $G\boldsymbol{\alpha}$ respectively.) However we will use $\boldsymbol{\alpha}$ for $\boldsymbol{\alpha}$ gauge transformed for ease of notation in what follows. (This sublety is not vital to the argument and must not be allowed to confuse us.) Since $A_0 = 0$ in the original gauge, the change in the original $A_0$ produced by $\boldsymbol{\alpha}$ (from (9.26)) is

$$\partial_0 \boldsymbol{\alpha}\,. \tag{9.28}$$

The matrix (9.23) is to be considered as the product of the two factors

$$\frac{\Delta(\partial_\mu A_\mu)}{\Delta(\boldsymbol{\alpha})} \quad \text{and} \quad \frac{\Delta(\boldsymbol{\alpha})}{\Delta(A_0)}. \tag{9.29}$$

So we have really factored the matrix which described how one gauge varies



with respect to the other gauge into two parts, each of which tells us how one of the gauges varies with respect to the infinitesimal gauge transformation $\boldsymbol{\alpha}$. The second of these two factors is independent of the fields since $\Delta(A_0)$ is independent of the fields (9.28). Using maths trick 1 we can drop this factor. The change in $\partial_\mu A_\mu$, is from (9.27):

$$\partial_\mu \boldsymbol{\alpha} \times A_\mu + \partial_\mu \partial_\mu \boldsymbol{\alpha}\,, \tag{9.30}$$

where we have used the fact that $\partial_\mu A_\mu = 0$ and we have put $\boldsymbol{\beta} = \boldsymbol{\alpha}$. This depends on the field $A_\mu$ and so we cannot get rid of the $\Delta(\partial_\mu A_\mu)/\Delta(\boldsymbol{\alpha})$ term in the same way as we got rid of $\Delta(A_0)/\Delta(\boldsymbol{\alpha})$.

We can now write (9.8) as

$$\int \exp\left\{ i \int [-\tfrac{1}{4} E_{\mu\nu} \cdot E_{\mu\nu} + J_\mu \cdot A_\mu]\, d^4x \right\} \delta(\partial_\mu A_\mu \mathcal{F}(A_\mu) \mathcal{D} A_\mu \mathcal{D}_{\text{Matter}}\,, \tag{9.31}$$

where $\mathcal{F}(A_\mu)$ is the Jacobian for the transformation from $\delta(A_0)$ to $\delta(\partial_\mu A_\mu)$ which we have identified as

$$\text{Det}\ \frac{\Delta(\partial_\mu A_\mu)}{\Delta(\boldsymbol{\alpha})}\,. \tag{9.32}$$

This can be written in the form

$$\int \exp\left\{ \frac{i}{2} \int P^+ (\partial_\mu \partial_\mu - A_\mu \times \partial_\mu) P\, d^4x \right\} \mathcal{D}P\,, \tag{9.33}$$

where $P$ is a Fermi particle. We can show this as follows. Let (9.32) be symbolically Det $c_{ij}$; this can be written, using maths trick 3 (9.21) as

$$\int \exp(\tfrac{1}{2} P_i^+ c_{ij} P_j) \prod_i dP_i\,. \tag{9.34}$$

Use (9.23) to re-write the exponent of this, giving

$$\int \exp\{-\tfrac{1}{2} P_i^* [y_i(x_j + P_j) - y_i(x_j)]\} \prod_i dP_i\,. \tag{9.35}$$

We now convert back from the symbolism to the $A$'s, putting as before $y_i \to \partial_\mu A_\mu$ and $x_i \leftrightarrow A_0$:

$$\int \exp\{-\tfrac{1}{2} P^* \Delta(\partial_\mu A_\mu)\} \mathcal{D}P\,. \tag{9.36}$$

Substituting for $\Delta(\partial_\mu A_\mu)$ from (9.30) and carrying out a Wick rotation, we get eq. (9.33). Integrating by parts in (9.33) and noting that $\partial_\mu A_\mu = 0$ by the choice of gauge, we get

$$\int \exp\left\{\frac{i}{2}\int [\partial_\mu P^{+} \cdot \partial_\mu P + \partial_\mu P^{+} \cdot (A_\mu \times P)]\,d^4x\right\} \mathcal{D}P\,, \tag{9.37}$$

and thus we have obtained the same propagator and coupling as we did in sect. 7.

In general, if we wish to transform to the gauge $f(A_\mu) = 0$, we first calculate

$$f(A_\mu + D_\mu \boldsymbol{\alpha}) = f(A_\mu) + f'(A_\mu) D_\mu \boldsymbol{\alpha}\,, \tag{9.38}$$

cf. (9.27) for the case $f(A_\mu) = \partial_\mu A_\mu$. Going through the above procedure

$$\delta(A_0) \to \delta\,[f(A_\mu)]\, \exp\left\{\frac{i}{2}\int P^{+} [f'(A_\mu) D_\mu] P\, d^4x\right\} \mathcal{D}P\,. \tag{9.39}$$

To summarise, we have transformed the path integral (9.7) to

$$\int \exp\left\{i \int [-\tfrac{1}{4} E_{\mu\nu} \cdot E_{\mu\nu} + J_\mu \cdot A_\mu + \tfrac{1}{2} \partial_\mu P^{+} \cdot \partial_\mu P + \tfrac{1}{2} \partial_\mu P^{+} \cdot (A_\mu \times P)]\, d^4x\right\}$$
$$\times\, \delta(\partial_\mu A_\mu) \mathcal{D}A_\mu \mathcal{D}P \mathcal{D}_{\text{Matter}}\,. \tag{9.40}$$

People have tried to make rules from this, but there is an easier way. From (9.14), we note that the determinant is independent of the value of $y_i(0)$ in the $\delta$-function. Hence if instead of choosing the gauge $\partial_\mu A_\mu = 0$ (i.e. $y_i(0) = 0$) we chose the gauge

$$\partial_\mu A_\mu(\underset{\sim}{x}, t) = \boldsymbol{\beta}(\underset{\sim}{x}, t) \quad \text{(i.e., } y_i(0) = \beta)\,, \tag{9.41}$$

the only changes this will produce will be to replace $\delta(\partial_\mu A_\mu)$ in (9.40) by $\delta(\partial_\mu A_\mu - \boldsymbol{\beta})$, and add an additional piece to 9.30. The last piece produces an additional term

$$\tfrac{1}{2} P^{+} \cdot (P \times \boldsymbol{\beta}) \tag{9.42}$$

which will cancel the extra term appearing in 9.37 when we integrate by parts since now $\partial_\mu A_\mu = \boldsymbol{\beta}$. We now that we can multiply 9.48 by the constant

$$\exp\left(-\frac{i}{2}\int \boldsymbol{\beta}^2\right) \mathcal{D}\boldsymbol{\beta}\,, \tag{9.43}$$

as this does not produce any difference in the theory.

Integrating over $\boldsymbol{\beta}$ using $\delta(\partial_\mu A_\mu - \boldsymbol{\beta})$ we obtain the following path integral

$$\int \exp\left\{i \int [-\tfrac{1}{4} E_{\mu\nu} \cdot E_{\mu\nu} - \tfrac{1}{2} (\partial_\mu A_\mu) \cdot (\partial_\nu A_\nu) + \tfrac{1}{2} \partial_\mu P^{+} \cdot \partial_\mu P\right.$$
$$\left. + \tfrac{1}{2} \partial_\mu P^{+} \cdot (A_\mu \times P)]\, d^4x\right\} \mathcal{D}A_\mu \mathcal{D}P \mathcal{D}_{\text{Matter}}\,. \tag{9.44}$$



We have recovered the Lagrangian (7.33) which led to the rules in tl gauge, and we have therefore shown that the gauges $A_0 = 0$ and $\partial_\mu$ equivalent.

It is possible to choose other gauges by replacing the term $\frac{1}{2}(\partial_\mu A$ by $(\lambda/2)(\partial_\mu A_\mu) \cdot (\partial_\nu A_\nu)$, where $\lambda$ takes any value between 0 and 1. course will produce a different effective Lagrangian. (What people r when they choose a gauge is to construct an effective Lagrangian an culate with it.) This produces a propagator for the $A_\mu$ of the form

$$\frac{\delta_{\mu\nu} - (1-\lambda)(k_\mu k_\nu / k^2)}{k^2}$$

which of course must produce the same physics. The modified prop useful when it comes to proving the renormalisability of gauge theo which I would have said more if I had more time) since if we take tl gauge we can then use naive power counting (i.e. the smart thing to power counting tells us whether something has a danger of divergin; doesn't prove anything: theories which look divergent by naive pow ing may not be, due to cancelations among the divergences. In som naive power counting makes the theory look horribly divergent but that this cannot be the case. In the $\lambda = 1$ gauge, there are no power: menta in the numerator of the propagator and so naively the power is better.

Fadde'ev and Popov, in fact, proceeded in the following way, with ing a specific gauge. Blindly if we were to write the simple simplest path integral for the theory, it would be

$$\int \exp\left\{i \int [-\tfrac{1}{4} E_{\mu\nu} \cdot E_{\mu\nu} + J_\mu \cdot A_\mu]\, d^4x\right\} \mathcal{D}A_\mu \mathcal{D}_{\text{Matter}}\,.$$

However, if we make a gauge transformation on the exponent, it dc change; so when we integrate over the $A_\mu$'s which are connected by transformation, we will get infinity. An analogy is when we evaluate t

$$\int \exp\{-\alpha(x^2 + 2xy + y^2)\}\, dx\, dy$$

Changing variables to $x - y$ and $x + y$, this becomes, up to a consta

$$\int \exp\{-\alpha(x+y)^2\}\, d(x+y)\, d(x-y)\,.$$

Since the exponent is independent of $(x - y)$, the integral over $(x$ - A similar thing happens in the path integral except that the situatic complicated: the Jacobian in the change of variables depends on $A$ is what introduces the ghosts we have been discussing.

Sirlin, Nucl. Phys.



## References

[1] C.N. Yang and R. Mills, Phys. Rev. 96 (1954) 191.
[2] J. Goldstone, Nuovo Cimento 19 (1961) 15.
[3] P.W. Higgs, Phys. Rev. Letters 12 (1964) 132.
[4] P.W. Higgs, Phys. Rev. Letters 13 (1964) 508.
[5] T.W. Kibble, Phys. Rev. 155 (1967) 1554.
[6] C.H. Llewellyn Smith, Phys. Letters B46 (1973) 233.
[7] S. Weinberg, Phys. Rev. Letters 19 (1967) 1264.
[8] A. Salam, in: Elementary Particle Theory, ed. N. Svartholm (Almquist and Forlag, Stockholm, 1968).
[9] J.D. Bjorken and S.D. Drell, Relativistic Quantum Mechanics (McGraw-Hill, 1968).
[10] E.S. Abers and B.W. Lee, Phys. Reports C9 (1973) 1.
[11] R.P. Feynman and A.R. Hibbs, Quantum Mechanics and Path Integrals (McGraw-Hill, 1965).
[12] L. Susskind, this volume, p. 207.
[13] R.P. Feynman, M. Kislinger and F. Ravndal, Phys. Rev. D3 (1971) 2706.
[14] P.W. Higgs, Phys. Rev. 155 (1966) 1554.
[15] B. Renner, Current Algebras and Their Applications (Pergamon, 1968).
[16] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman Lectures in Physics, Vol. 3 (1965).
[17] D. Rogovin and M. Scully, Phys. Reports 25 (1976) 175.
[18] R.P. Feynman, Rev. Mod. Phys. 20 (1948) 367.
[19] C. DeWitt-Morette, private communication (1951); General Relativity and Gravitation Journal, to be published.
[20] C. Garrod, Rev. Mod. Phys. 38 (1966) 483.
[21] P.A.M. Dirac, Principles of Quantum Mechanics (Oxford University Press, 1957).
[22] J. Schwinger, in: Theoretical Physics, Trieste Lectures, 1962 (IAEA, Vienna, 1963).
[23] R.P. Feynman, Acta Phys. Polonica 26 (1963) 697.
[24] B.S. DeWitt, Phys. Rev. 162 (1967) 1195, 1239.
[25] G. 't Hooft, Nucl. Phys. B35 (1971) 167.
[26] L.D. Fadde'ev and V.N. Popov, Phys. Letters B25 (1967) 29.
[27] V.N. Popov and L.D. Fadde'ev, Perturbation Theory for Gauge Invariant Fields, Kiev I.T.P. Report (unpublished).