WHY DO WE NEED LOCAL GAUGE INVARIANCE IN THEORIES WITH VECTOR PARTICLES? AN INTRODUCTION

G. 't Hooft

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

During the last decennium it has become more and more clear that gauge theories play a decisive role in elementary particle physics. Indeed, the most general possible theory for a finite class of particle types, with not too strong interactions, and valid to very high energies, *must* be a gauge theory with some scalar and spinor fields.

Renormalizable field theories with only scalar and spinor fields with Yukawa type interactions were known for a long time. How to introduce vector fields has for a long time been an outstanding problem, which was finally solved with the advent of the gauge theories. In this lecture I will give field theory in a nutshell, starting by "defining" functional integrals, applied to scalar field theories. Then I shall discuss the so-called cutting-relations. These crucial relations in this form were derived by M. Veltman in his pioneering work in a time that few believed in field theory and nobody in gauge theories. Finally we will see how unitarity and renormalizability can only be reconciled in a gauge theory.

I1. THE PRECURSORS OF FUNCTIONAL INTEGRALS

Consider integrals of the type:

$$Z(\vec{J},\vec{\lambda}) = \int \dots \int d\phi_1 \dots d\phi_n \exp\left[-\tfrac{1}{2}\,\phi_i\, M_{ij}\, \phi_j - \lambda_1 \sum_i \phi_i^3 - \lambda_2 \sum_i \phi_i^4 \dots + J_i\, \phi_i\right] . \tag{I1}$$

Here $M_{ij} = M_{ji}$, and the coefficients M_{ij} and λ_i may either be real or imaginary with a small positive real part. Let us expand with respect to $\lambda_{1,2}$ and J_i:

$$Z(\vec{J},\vec{\lambda}) = \int \ldots \int \prod_i d\phi_i \; e^{-\frac{1}{2}(\phi,M\phi)} \Big[1 - \lambda_1 \sum \phi_i^3 \ldots$$
$$\ldots + \tfrac{1}{2} \sum J_i J_j \phi_i \phi_j \ldots - \frac{1}{3!} \sum J_i J_j J_k \phi_i \phi_j \phi_k \lambda_1 \sum_\ell \phi_\ell^3 \text{ etc.} \Big] . \quad \text{(I2)}$$

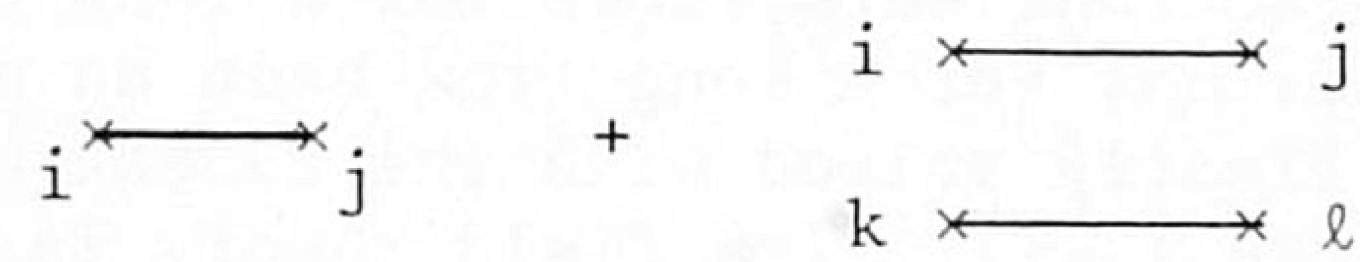
The integrals can now be performed, for instance by diagonalizing M. But more conveniently, one can write

$$\prod_i \int d\phi_i \; \phi_i \phi_j \phi_k \ldots e^{-\frac{1}{2}(\phi,M\phi)} =$$
$$= \frac{\partial}{\partial J_i} \frac{\partial}{\partial J_j} \frac{\partial}{\partial J_k} \ldots \prod_i \int d\phi_i \; e^{-\frac{1}{2}(\phi,M\phi) + (J,\phi)} \Big|_{\vec{J}=0} =$$
$$= \frac{\partial}{\partial J_i} \frac{\partial}{\partial J_j} \frac{\partial}{\partial J_k} e^{+\frac{1}{2}(J,M^{-1}J)} \; Z(o,o) \; . \quad \text{(I3)}$$

And it is easy to expand

$$e^{\frac{1}{2}(J,M^{-1}J)} = 1 + \tfrac{1}{2} J_i (M^{-1})_{ij} J_j + \frac{1}{2!4} J_i (M^{-1})_{ij} J_j J_k (M^{-1})_{k\ell} J_\ell \ldots . \quad \text{(I4)}$$

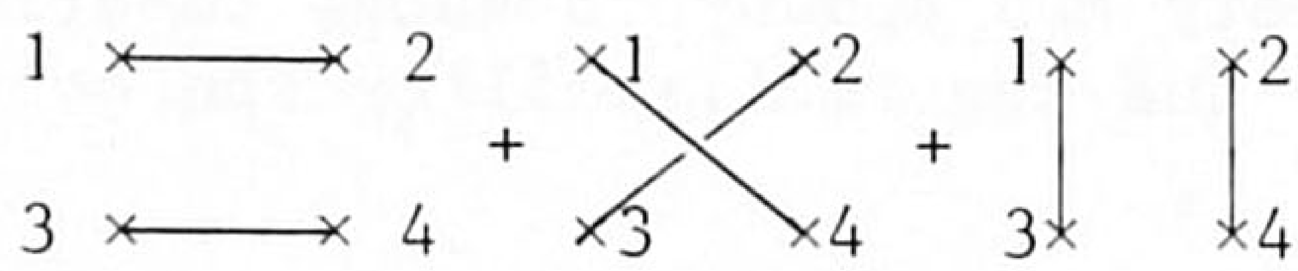
The resulting expressions are conveniently expressed in terms of diagrams. The right hand side of (I4) is to be written as

1 + 

+ ...

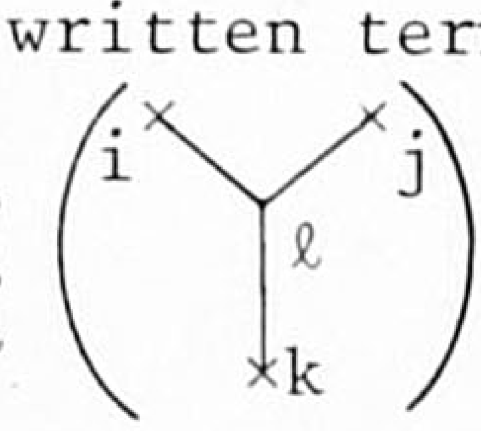
The factors $\frac{1}{2}$,... can often be incorporated in the diagrams. For instance one could write the third term of eq. (I4) by differentiating with respect to four different J terms:

$$\frac{\partial}{\partial J_1} \frac{\partial}{\partial J_2} \frac{\partial}{\partial J_3} \frac{\partial}{\partial J_4} (\ldots) =$$



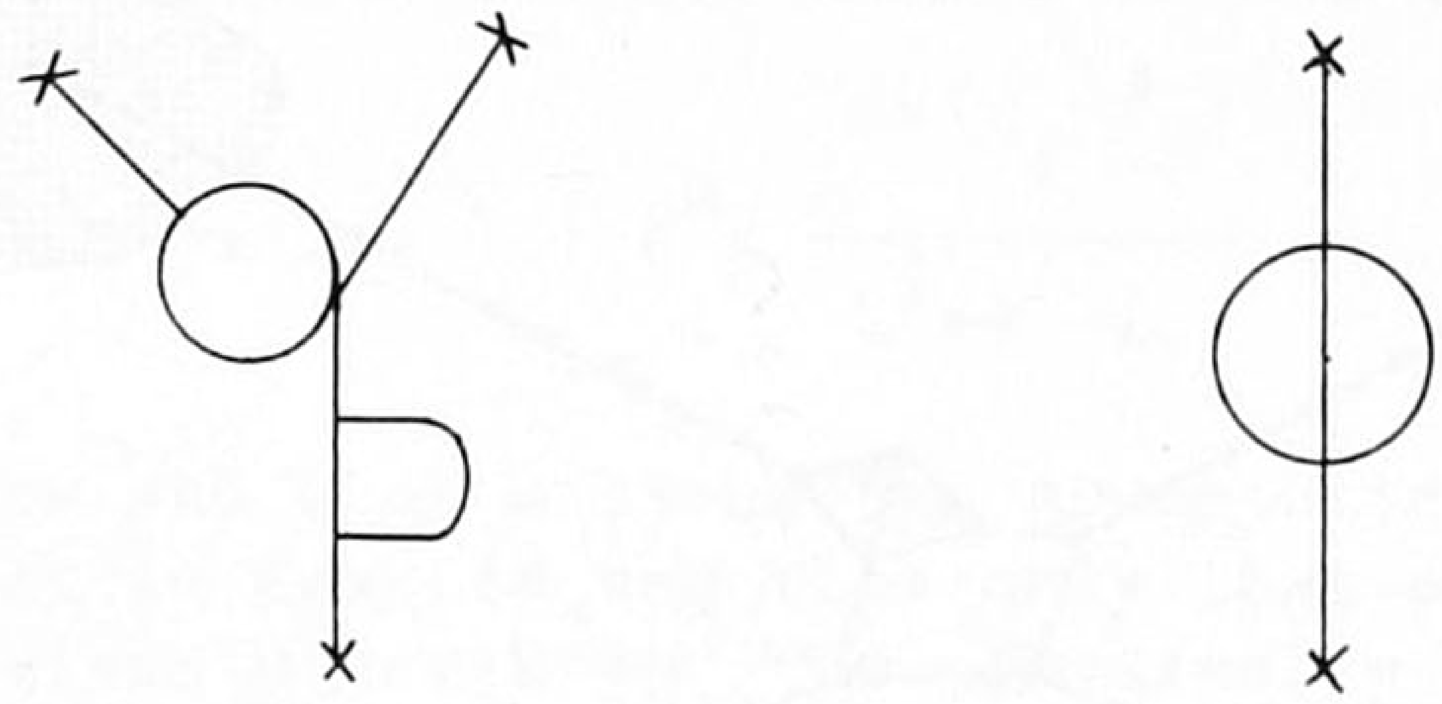
$$= (M^{-1})_{12} (M^{-1})_{34} + (M^{-1})_{14} (M^{-1})_{23} + (M^{-1})_{13} (M^{-1})_{24} \; . \quad \text{(I5)}$$

It is now easy to compute the terms in (I2). For instance the last explicitly written term:

$$-\frac{\lambda_1}{3!} \sum_\ell$$ 

$$= -\frac{\lambda_1}{3!} J_i J_j J_k \sum_\ell (M^{-1})_{i\ell} (M^{-1})_{j\ell} (M^{-1})_{k\ell} \; .$$

Still further, one gets diagrams such as



If one writes

$$Z(\vec{J},\vec{\lambda}) = Z(o,o)e^{W(\vec{J},\vec{\lambda})} \quad , \tag{I6}$$

then W contains only *connected* diagrams. Further procedures are possible to obtain only *irreducible* diagrams, but we will not dwell any further on that. So far, exact and obvious mathematics.

I2. SCALAR FIELD THEORIES

In a scalar field theory the variables in the integrals to be considered are not a finite but an infinite set, namely the values of the field ϕ at each space-time point x. So we formally replace in the previous section

$$\phi_i \quad \text{by} \quad \phi(x) \ ,$$

where x is a vector in $\mathbb{R}^n$. To be explicit we consider $\mathbb{R}^n$ to be Minkowsky space. Usually we have

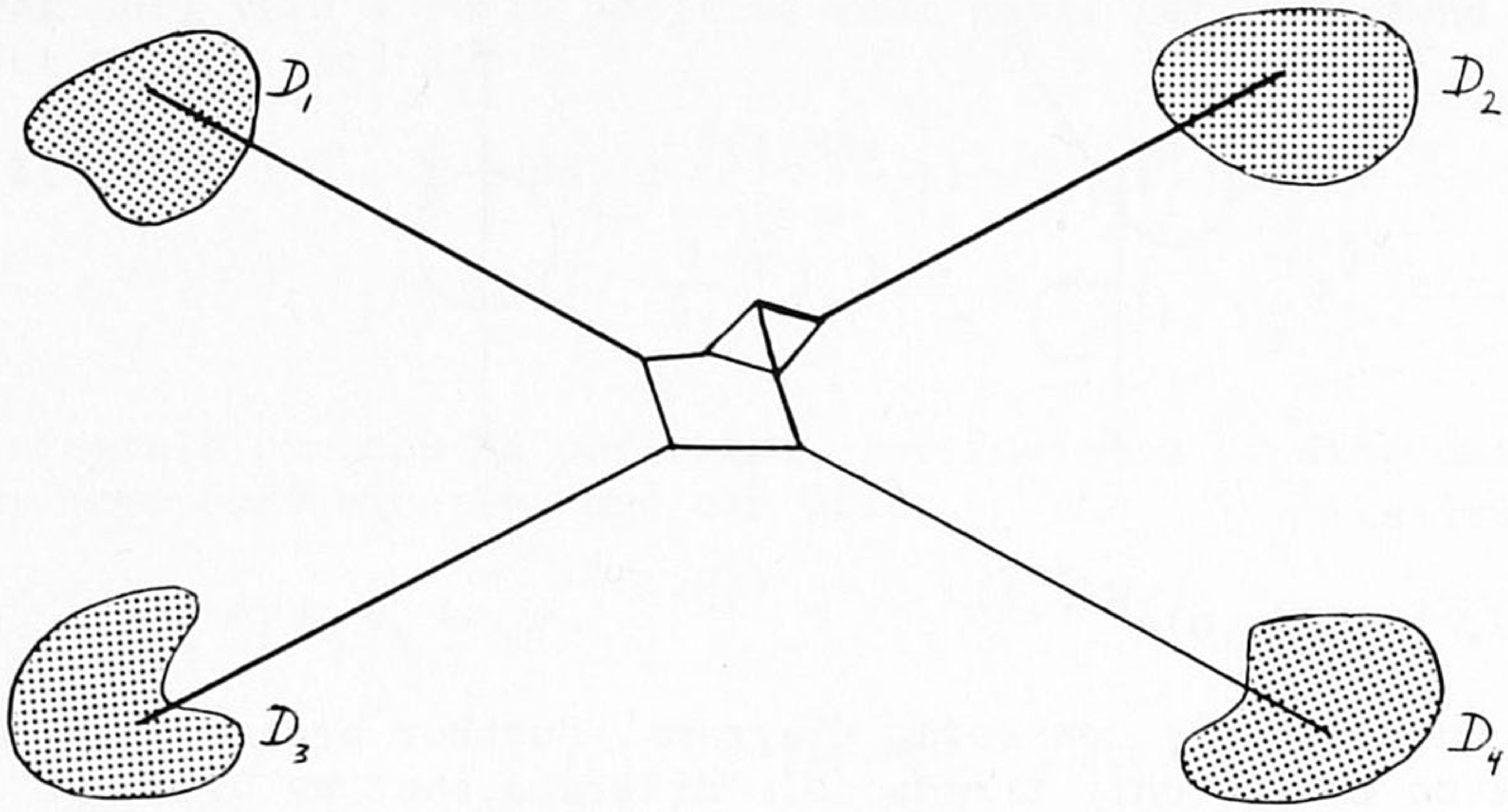
$$M = i\left(-\partial_x^2 + m^2\right) \ . \tag{I7}$$

The integral becomes an "integral" over functions $\phi(x)$:

$$Z\{J,\vec{\lambda}\} = \int D\phi \exp i\left[\int d^n x\left(-\tfrac{1}{2}\phi(x)(m^2-\partial^2)\phi(x) - \lambda_1\phi^3(x) - \lambda_2\phi^4(x) + \right.\right.$$
$$\left.\left. + J(x)\phi(x)\right)\right] \ , \tag{I8}$$

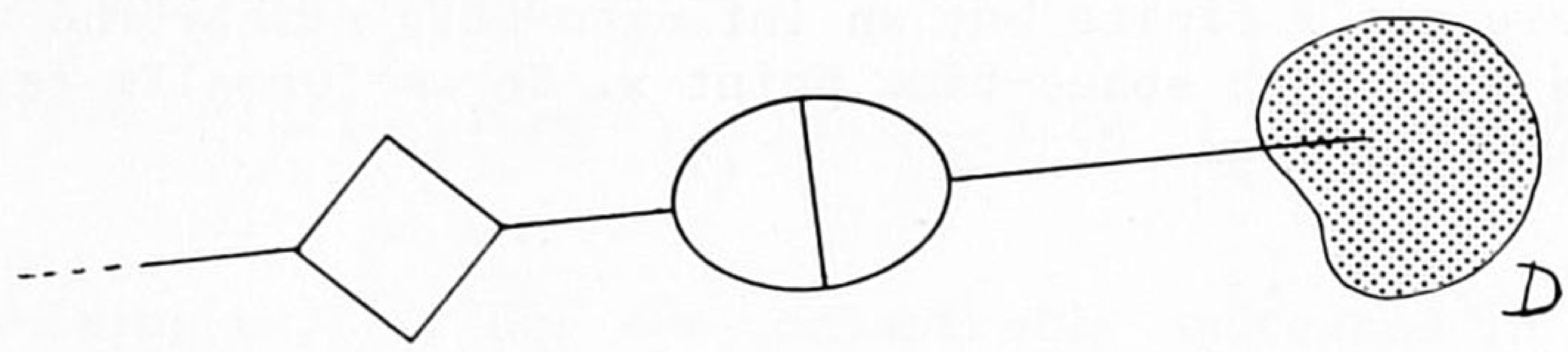
a functional integral which, particularly for $n > 3$, is devoid of any rigorous meaning. If $n = 4$ and only λ_1 and λ_2 are different from zero then our expressions can be suitably altered in such a way that they make sense. But it is not easy. We restrict ourselves always to perturbation expansions, which implies that we only consider the Feynman graphs, not the entire integral.

Let us consider $Z(J,\lambda)$ as a functional of the function $J(x)$. Let us take $J(x) = \sum_\ell J_\ell(x)$ where $J_\ell(x)$ are functions with compact support D_ℓ and the domains D_ℓ move away from each other to infinity. Pictorially, we have the diagrams



(I9)

Let us, for simplicity, ignore blobs that may occur in the external lines:



(I10)

(these can easily be treated later).
It is convenient to diagonalize the matrices M, which brings us to the space of Fourier transforms of $\phi(x)$, namely $\{\phi(k)\}$. In this space we have

$$M^{-1} = \frac{-i}{k^2+m^2} .$$

Now in Minkowsky space $k^2 + m^2$ can be zero. We improve the definition of our functional integral by replacing $k^2 + m^2$ by $k^2 + m^2 - i\varepsilon$, so that

$$e^{-\phi(-\partial^2+m^2)\phi}$$

is replaced by

$$e^{-i\phi(-\partial^2+m^2)\phi - \varepsilon\phi^2} , \qquad \text{(I11)}$$

which converges if $-\partial^2+m^2$ happens to vanish.
Of course, ε is positive infinitesimal. We now have in x-space

$$M^{-1}(x_1,x_2) = -i(2\pi)^{-n} \int d^nk \, \frac{e^{ik(x_1-x_2)}}{k^2+m^2-i\varepsilon} \, . \tag{I12}$$

One may now ask what will be the asymptotic form of M^{-1} as $|x_1-x_2| \to \infty$. It is easy to see that only that part of the integral (I11) survives where $k^2+m^2 \simeq o$. Physically that implies that the particles go to their mass shells. The sources J_ℓ at the domains D_ℓ can physically be interpreted as particle production or detection machines, far away from the interaction region. The amplitude approaches the scattering matrix S. We now short-circuit complicated lengthy definitions of S, defining it simply directly from the Feynman diagrams.

1) Consider all diagrams of the type I9 and replace in the external lines

$$\frac{1}{k^2+m^2-i\varepsilon} \quad \text{by} \quad \delta(k^2+m^2) \, , \tag{I13}$$

that is, we only consider external lines on mass shell, in the Fourier picture.

2) Look at the sign of $k_o = \pm\sqrt{m^2+\vec{k}^2}$. If $k_o > o$ we interpret this as an outgoing particle, if $k_o < o$ it is an ingoing particle.

3) The diagram is now a contribution to the matrix element

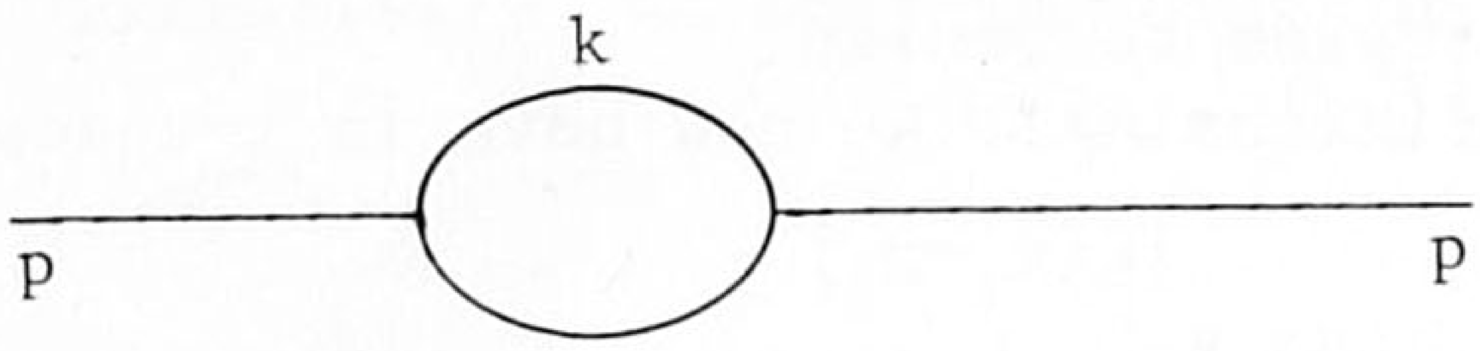
$$_{\text{outgoing}}\langle k_1,k_2,\ldots|S|k_3,k_4\ldots\rangle_{\text{ingoing}} \, . \tag{I14}$$

The normalization is to be fixed later.

Formally, this all can be derived from good old quantum mechanics by writing the evolution kernel

$$U(t_1,t_2) = e^{-i(t_1-t_2)H}$$

as a path integral, for the case of a set of anharmonic oscillators at all points x. This is why one expects that the S matrix defined this way will come out to be unitary. However, divergences, for instance in the diagram

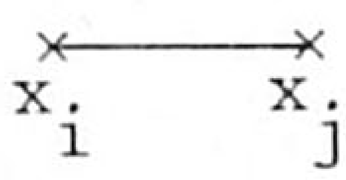


$$\int \frac{d^4k}{(k^2+m^2-i\varepsilon)((k+p)^2+m^2-i\varepsilon)} \tag{I15}$$

make redefinitions necessary. We must require that such redefinitions do not affect unitarity. Furthermore, we will want to write down functional integrals for the S matrix elements for particles with spin, in which case the canonical method is less direct. In short, we wish to understand explicitly how unitarity follows directly from the diagrams.

I3. CUTTING RELATIONS (SIMPLIFIED)

Consider the "propagator" in Fourier space



$$\frac{1}{(2\pi)^4 i}\,\frac{1}{k^2+m^2-i\varepsilon} \equiv \Delta(k)\ . \tag{I16a}$$

In x-space:

$$\Delta(x_i,x_j) = \int d^4k\ e^{ik(x_i-x_j)}\ \Delta(k) \equiv \Delta_{ij}\ . \tag{I16b}$$

Here, i and j are labels for two vertices in a diagram. x_i and x_j are the corresponding x variables. We write.

$$\Delta_{ij} = \theta(x_o)\Delta^+_{ij} + \theta(-x_o)\Delta^-_{ij}\ , \tag{I17}$$

$$\Delta^{\pm}_{ij} = \frac{1}{(2\pi)^3} \int d_4k\ e^{ikx}\ \theta(\pm k_o)\ \delta(k^2+m^2)\ , \tag{I18}$$

where $x = x_i - x_j$.
We have

$$\Delta^{\pm}_{ij} = (\Delta^{\mp}_{ij})^* = \Delta^{\mp}_{ji}\ . \tag{I19}$$

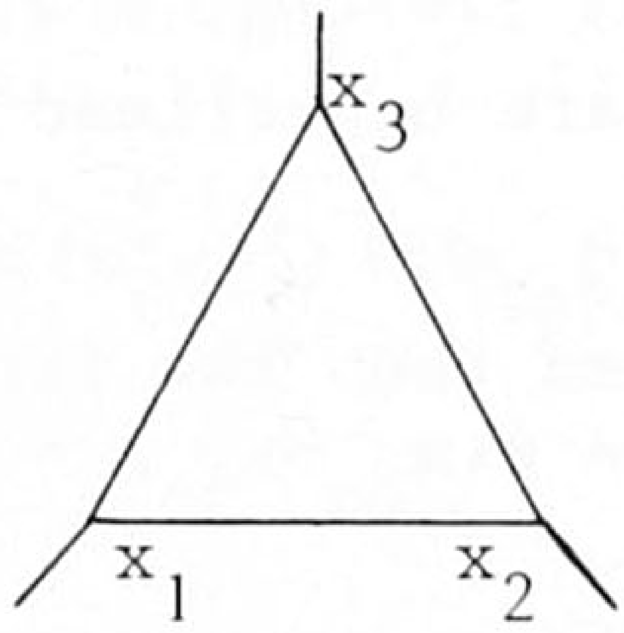
Therefore,

$$\Delta^*_{ij} = \theta(x_o)\Delta^-_{ij} + \theta(-x_o)\Delta^+_{ij}\ . \tag{I20}$$

The above identities are easily derived from

$$\theta(x) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} d\tau\ \frac{e^{i\tau x}}{\tau-i\varepsilon}\ . \tag{I21}$$

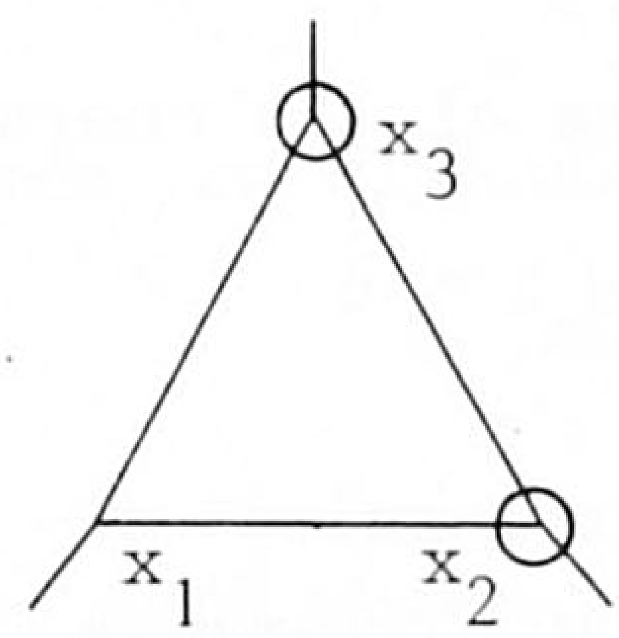
Consider now a diagram in the x-representation



leaving out the δ functions for the external lines. Call it $F(x_1, x_2, x_3, \ldots)$ (here we sketched the diagram for $F(x_1, x_2, x_3) = (-i\lambda_1)(-i\lambda_2)(-i\lambda_3)\,\Delta_{12}\,\Delta_{23}\,\Delta_{31}$). This diagram may contribute to the S matrix after Fourier transformation. Now we define a more general class of diagrams

$$F(x_1, \ldots, \underline{x}_i, \ldots, \underline{x}_j, \ldots x_n) ,$$

with some of the arguments x underlined. In the diagram we denote this by drawing a circle around those vertices:



This function F is defined by the following replacements: replace Δ_{ij} by

1) Δ_{ij} if neither x_i nor x_j are underlined,

2) Δ^+_{ij} if x_i but not x_j is underlined,

3) Δ^-_{ij} if x_j but not x_i is underlined, (I22)

4) Δ^*_{ij} if both x_i and x_j are underlined,

5) Replace one factor i by $-i$ for every underlined x (remember they carry a factor $-i\lambda$; replace by $+i\lambda$, not $i\lambda^*$).

In the above example we have

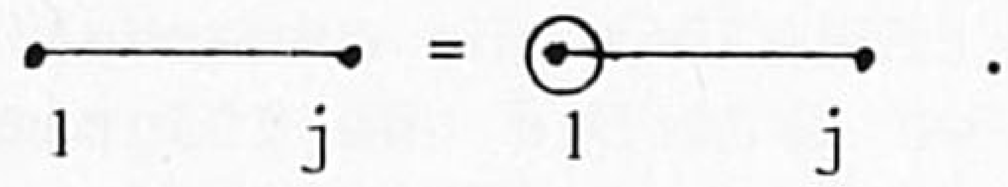
$$F(x_1, \underline{x}_2, \underline{x}_3) = (-i\lambda_1)(i\lambda_2)(i\lambda_3)\,\Delta^-_{12}\,\Delta^*_{23}\,\Delta^+_{31} .$$

We now have <u>theorem 1</u> ("largest time equation"). Let x_1 be the x-coordinate with largest time component, $x_{1o} > x_{io}$ for all i. Then

$$F(x_1,\ldots) + F(\underline{x}_1,\ldots) = o \ , \tag{I23}$$

where all other coordinates are underlined or not in the same way for both terms.

Proof: if x_j is not underlined then the first term contains Δ_{1j}, the second Δ^+_{1j}. But because $\theta(x_{1o}-x_{jo}) = 1$ we have (see eq. I17):

$\Delta_{1j} = \Delta^+_{1j}$, or  .

Similarly, if x_j is underlined we have Δ^-_{1j} in the first and Δ^*_{ij} in the second term. Because of eq. (I20),

$\Delta^-_{1j} = \Delta^*_{1j}$, or  .

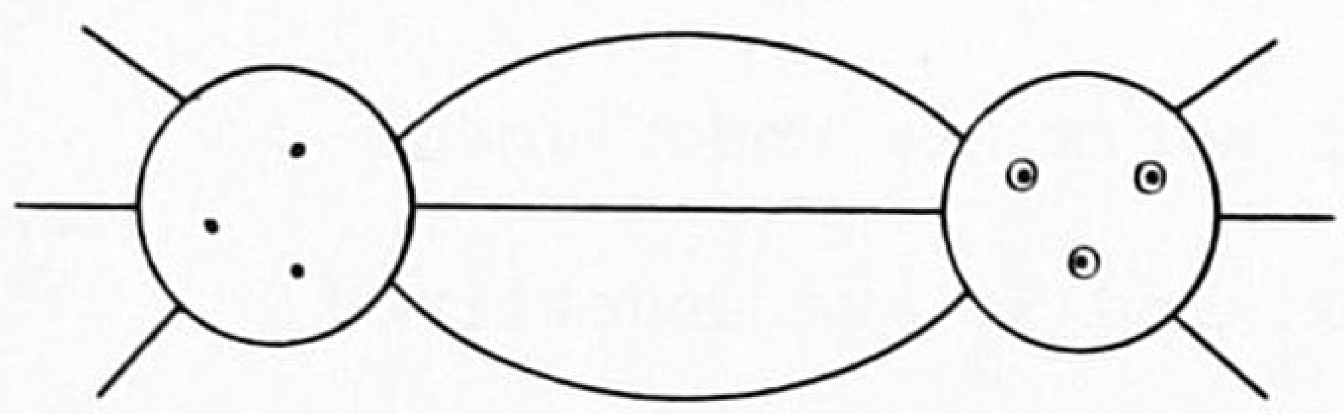
Because of replacement # 5 we have one sign flip, from which follows (I23).

<u>Theorem 2</u>. For all values of the coordinates x_i we have

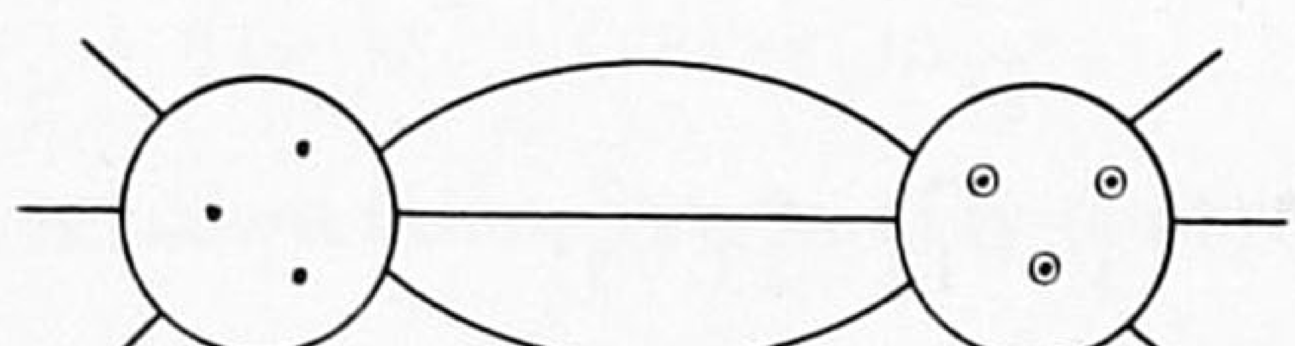
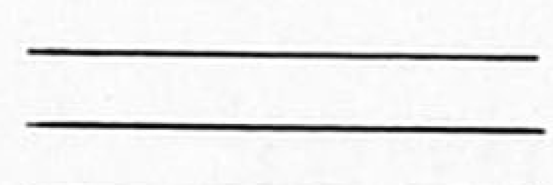
$$\sum_{\substack{\text{all possible}\\ \text{underlinings}}} F(x_1,\ldots, x_n) = o \tag{I24}$$

Proof: There is always one x with largest time component. The diagrams therefore combine in pairs of terms that cancel each other.

Next, let us deform all terms in (I24) in such a way that the underlined vertices occur at the right, the non-underlined at the left. We obtain:

$\sum$  = o .

Add the trivial diagram (without any vertices) and sum formally over all diagrams. We see

$\sum$  = $\sum$  .(I25)

How do we interprete this expression? Let the diagrams with all vertices encircled (all coordinates underlined) generate a matrix

$\tilde{S}$. The lines connecting S with $\tilde{S}$ are on mass shell with positive energy component k , because they all correspond to a term Δ^+_{ij}, see (I18). Our eq. (I25) says

$$\int dk_i <A|S|k_1 k_2 k_3,\ldots> \prod_i \delta(k_i^2+m_i^2)\,\theta(k_{io})\, <k_1 k_2 k_3,\ldots|\tilde{S}|B> = <A|B>, \tag{I26}$$

which is unitarity if $\tilde{S} = S^\dagger$. (I27)

One easily checks that indeed the diagrams for $\tilde{S}$ are the same as the ones for $S^\dagger$ provided that the coupling constants λ in eq. (I8) are real. Otherwise the replacement # 5 in (I22) would not imply complex conjungation.

What generalizations are allowed such that unitarity is not lost? Suppose our propagator was a matrix of the form

$$\frac{\rho_{ij}}{k^2+m^2-i\varepsilon}, \tag{I28}$$

then the intermediate lines (in I25) would carry an extra matrix ρ:

$$S\,\rho\,S^\dagger = I\,. \tag{I29}$$

We reobtain unitarity by redefining the norm of the particle states with a factor $\sqrt{\rho}$, but this only works if

<u>ρ is positive definite</u>. (I30)

If ρ has zero eigenvalues besides positive ones then one can write

$$\rho = \sum_i |\psi_i> \rho_i <\psi_i|, \tag{I31}$$

where the number of states is less than the dimension of ρ. That describes a system with fewer types of particles than fields ϕ and is therefore acceptable.

I4. REGULATORS

One would like to change the theory further in such a way that integrals such as (I15) are made convergent (regularized). An example is the Pauli-Villars regulator. We replace all propagators $1/k^2+m^2-i\varepsilon$ by

$$\frac{1}{(k^2+m^2-i\varepsilon)(1+\frac{k^2}{\Lambda^2}-i\varepsilon)} = \frac{\Lambda^2}{\Lambda^2-m^2}\left(\frac{1}{k^2+m^2-i\varepsilon} - \frac{1}{k^2+\Lambda^2-i\varepsilon}\right) \tag{I32}$$

We immediately see that the minus sign destroys unitarity. A fictitious particle with mass Λ may be produced, but in view of the sign of its contribution the production "probability" for any odd number of these objects is negative. However we also observe that our identities can still be used. An immediate Λ line in (I25) is now associated with a factor

$$\delta(k^2+\Lambda^2)\ \theta(k_o)$$

and since $k_o = \pm\sqrt{\vec{k}^2+\Lambda^2}$ this only contributes if $k_o > \Lambda$. All other energies k_o in the intermediate lines are positive as well. Therefore, we find only violation of unitarity in channels where the total energy in the initial and final states $\langle A|$ and $|B\rangle$ exceeds Λ.
We can now take the limit $\Lambda \to \infty$ carefully, inserting explicitly Λ dependent extra terms ("counter terms") in the Lagrangian. A theory is called "renormalizable" if such a limit can be taken in such a way that the S matrix according to our definition remains finite. For scalar field theories this requires the absence of any couplings other than cubic or quartic in the fields ϕ.

I5. PARTICLES WITH SPIN ONE

We skip the case of spin $\frac{1}{2}$ particles which yields the well-known Dirac equation and Fermi-Dirac statistics. We wish to construct the theory for particles with spin one. Let us try to construct the propagator, first when the particle is at rest: $k = (k_o,o,o,o)$. (Relativistic covariance will give us the moving case.) The states $|\psi_i\rangle$ in (I31) must transform as vectors under the "little group" rotations (these are the SO(3) rotations that leave the form-vector k_μ invariant). Suppose that we choose for the matrix $\rho_{\mu\nu}$ in (I28)

$$\rho_{\mu\nu} = g_{\mu\nu} = \text{diag}\ (-1,1,1,1)\ . \qquad \text{(I33)}$$

Then unitarity would clearly be violated. The minus sign in (I33) would correspond to an unwanted particle with negative metric. This is why the Lagrangian (= term in the exponent of the functional integral)

$$\mathcal{L} = -\tfrac{1}{2}(\partial_\mu A_\nu)^2 - \frac{m^2}{2} A_\mu^2\ , \qquad \text{(I34)}$$

for a massive vector particle is unacceptable. We want for the particle at rest

$$\rho_{\mu\nu}(k) = \text{diag}\ (o,1,1,1)\ . \qquad \text{(I35)}$$

For general k_μ that could be

$$\rho_{\mu\nu}(k) = g_{\mu\nu} - \frac{k_\mu k_\nu}{k^2 \pm i\varepsilon} . \tag{I36}$$

But, whatever sign we choose for $\pm i\varepsilon$, the extra part introduces an extra pole at $k^2 = o$ with wrong sign (the factors k_μ were not yet considered in my simplified cutting relations, but when they are carefully taken into account one verifies that the sign corresponds to particles with wrong metric).

The only correct Lorentz invariant propagator is

$$\frac{\delta_{\mu\nu} + k_\mu k_\nu / m^2}{k^2 + m^2 - i\varepsilon} . \tag{I37}$$

One verifies (I35) on mass shell. No extra spurious poles occur corresponding to unwanted particle states. The propagator has only one major disadvantage: it is badly divergent as $k \to \infty$. This could also imply complications in our unitarity relations as $x_i - x_j \to o$ because the θ functions have not yet been properly defined there. Suitable addition of regulators could cure such problems. However because of the $k_\mu k_\nu$ terms the limits $\Lambda \to \infty$ for these regulators are much harder to take. Indeed, in a general theory for vector particles with such propagators the desease of high-momentum infinities would spread beyond control. Note that the Lagrangian giving this propagator is

$$\mathcal{L} = -\tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 - \tfrac{1}{2}m^2 A_\mu^2$$
$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}^2 - \tfrac{1}{2}m^2 A_\mu^2 . \tag{I38}$$

I6. GAUGE TRANSFORMATIONS

We now wish to cure the infinity problems without destroying the good (= unitary) features of the Lagrangian (I38). Consider the functional integral

$$Z = \int DA_\mu \left(\det \frac{\partial C(A)}{\partial \Omega} \right) \exp i \int d^n x \left[\mathcal{L}^{\rm inv}(A) - \tfrac{1}{2}(C(A))^2 \right] . \tag{I39}$$

Here $\mathcal{L}^{\rm inv}$ is a function of $A_\mu(x)$ and $\partial_\mu A_\nu(x)$ which is invariant under local gauge transformations Ω of $A_\mu(x)$:

$$A'_\mu(x) = \Omega(x) \left[A_\mu(x) + \frac{1}{g}\partial_\mu \right] \Omega^{-1}(x) . \tag{I40}$$

Here $A_\mu(x)$ is written as an antihermitian matrix. $C(A)$ is a gauge-fixing term. That is, the restriction $C(A) = o$ can always be satisfied for any field configuration A after performing gauge rotations of the type (I40). Finally,

$$\det\left(\frac{\partial C(A)}{\partial \Omega}\right)$$

is a formal functional determinant of an operator defined by subjecting C to an infinitesimal gauge rotation Ω. In fact, this determinant simply fixes a particular measure (through still superficially) for the functional integral. Let us give an example. The infinitesimal gauge rotations are:

$$A_\mu^a(x) \rightarrow A_\mu^a(x) - \partial_\mu \Lambda^a(x) + g f_{abc}\, \Lambda^b(x)\, A_\mu^c(x) \; ; \qquad \text{(I41)}$$

$$C^a(x) = \partial_\mu A_\mu^a(x) \; , \qquad \text{(I42)}$$

$$\mathcal{L}^{inv}(A) = -\tfrac{1}{4}\, G_{\mu\nu}^a(x)\, G_{\mu\nu}^a(x) \; , \qquad \text{(I43)}$$

$$G_{\mu\nu}^a(x) = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f_{abc}\, A_\mu^b A_\nu^c \; , \qquad \text{(I44)}$$

$$\frac{\partial C^a(x)}{\partial \Omega^b(x)} = \delta_{ab}\partial_\mu^2 + g f_{abc}\, \partial_\mu (A_\mu^c) \; . \qquad \text{(I45)}$$

Here f_{abc} are structure constants of some compact group. The determinant is something new. But, we can give it a more familear appearance, by writing

$$\det{}^{-N} \frac{\partial C}{\partial \Omega} = \int D\phi D\phi^* \, e^{-\phi_i^* \frac{\partial C}{\partial \Omega} \phi_i} \; , \qquad \text{(I46)}$$

where ϕ_i have N components. In our case:

$$-\phi_i^* \frac{\partial C}{\partial \Omega} \phi_i = \int dx \left[-\partial_\mu \phi_i^{a*} \partial_\mu \phi_i^a - g f_{abc}\, \partial_\mu \phi_i^{a*}\, A_\mu^b \phi_i^c \right] \qquad \text{(I47)}$$

(where we performed a partial integration).
The Feynman rules for

$$\det \frac{\partial C}{\partial \Omega}$$

are now easily read off from this "Lagrangian" in the usual way. For the factor N, associated to each closed ϕ-loop, we must substitute −1.

Theorem: the function Z defined in (I39) is independent of the choice of the gauge fixing function C(A).
The proof of this theorem can be given in various ways, either by combinatorics with diagrams or by first proving the assertion for finite dimensional integrals. One of the basic ingredients of the original proofs of renormalizability of the gauge theories was to show that this theorem remains valid even if the expression (I39) for Z has been made finite by special regularization techniques (dimensional regularization). I will not go into this but simply

formulate how in principle this central theorem can be used to obtain a renormalizable theory.

I7. A GAUGE THEORY

Let us consider as a gauge group SU(2). Let the fields be A_μ^a and a complex doublet ξ_i.

$$\mathcal{L}^{inv}(A_\mu^a, \xi_i, \xi_i^*) = -\tfrac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a - D_\mu\xi^* D_\mu\xi$$

$$-\tfrac{1}{2}\lambda(\xi^*\xi - F^2)^2 \ . \qquad \text{(I48)}$$

Here F is a c-number, and

$$D_\mu\xi = \partial_\mu\xi - \tfrac{1}{2}ig\, A_\mu^a\, \tau^a\xi \ . \qquad \text{(I49)}$$

$\tau^{1,2,3}$ are the 2x2 Pauli matrices.
First choice of gauge (the so-called unitary gauge):

$$C^a(A,\xi) = \alpha \begin{pmatrix} \mathrm{Re}\ \xi_2 \\ \mathrm{Im}\ \xi_1 \\ \mathrm{Im}\ \xi_2 \end{pmatrix} \ ; \quad \alpha \to \infty \qquad \text{(I50)}$$

It is easily seen that if $\alpha \to \infty$ this simply corresponds to freezing out the variables

$$\mathrm{Re}\ \xi_2\ ,\ \mathrm{Im}\ \xi_1\ ,\ \mathrm{Im}\ \xi_2 \to 0 \ .$$

Let us write

$$\xi = \begin{pmatrix} F \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix} Z \\ 0 \end{pmatrix} . \qquad \text{(I51)}$$

where Z is now a single real scalar field. The functional determinant

$$\det \frac{\partial C}{\partial \Omega}$$

gets no space-time derivatives. This implies that the effects of this determinant, through highly divergent, vanish completely when we regularize and add counter terms (not explained any further here).
Substituting (I51) we obtain

$$\mathcal{L} = -\tfrac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a - \tfrac{1}{4}g^2F^2A_\mu^2 - \tfrac{1}{2}(\partial_\mu Z)^2 - \lambda F^2 Z^2$$

$$+ \text{ local interaction terms.} \qquad \text{(I52)}$$

Observe that the vector field A_μ^a occurs precisely in the combination that gives a unitary propagator (see I38). The mass is $gF/\sqrt{2}$. The scalar field Z has mass $F\sqrt{2\lambda}$. We have a unitary theory with three- spin-one and one spin-sero particle types.

What makes this theory so special? Let us consider a second choice of gauge (the so-called renormalizable gauge):

$$C^a(A,\xi) = \partial_\mu A^a_\mu - gF\begin{pmatrix} I_m\ \xi_2 \\ -\mathrm{Re}\ \xi_2 \\ \mathrm{Im}\ \xi_1 \end{pmatrix} , \qquad (I53)$$

a special combination chosen only for convenience. We get

$$\mathcal{L}^{inv} = -\tfrac{1}{2}(\partial_\mu A^a_\nu)^2 - \tfrac{1}{4}g^2F^2A^2_\mu -$$

$$-\tfrac{1}{2}(\partial_\mu \psi^a)^2 - \tfrac{1}{4}g^2F^2(\psi^a)^2 - \tfrac{1}{2}\partial_\mu Z^2 - \lambda F^2Z^2$$

$$+ \text{ total derivatives } + \text{ local interaction terms} \qquad (I45)$$

where

$$\xi_i = \begin{pmatrix} F \\ o \end{pmatrix} + \frac{1}{\sqrt{2}}\begin{bmatrix} i\psi^3+Z \\ -\psi^2+i\psi^1 \end{bmatrix} .$$

The total derivatives are irrelevant.
The vector field Lagrangian is now rather like (I34) and gives nicely convergent propagators:

$$\frac{g_{\mu\nu}}{k^2+\frac{1}{2}g^2F^2-i\varepsilon} . \qquad (I55)$$

Only due to gauge invariance the two theories are equivalent. The renormalizable gauge still shows infinities but they are not worse than in the scalar and spinor field theories. A renormalization procedure can remove them.

The g_{oo} part of the propagator (I55) describes a spurious particle or "ghost". Suppose we apply our unitarity criterion directly. We would notice

$$S\ K\ S^\dagger = I , \qquad (I56)$$

where K is an operator with eigenvalues ±1. K has a factor −1 for each of these spurious particles. Furthermore, the complex scalar field ϕ in (I47) now also contributes. It is produced in pairs. Because their multiplicity factor N is −1, they also come with wrong metric: a factor −1 for each ϕ, anti-ϕ pair. Finally, the "particles" described by the fields ϕ^a in (I54) have positive metric but should not be considered as real particles because they are absent in (I52). All these spurious states cancel in (I56). If we write Hilbert space as a product of a physical Hilbert space (containing only real particles) and a ghost space G (of all states with at least one unphysical particle), then

$$I = S K S^{\dagger} = \sum_{P} S|P\rangle\langle P|S^{\dagger} + \sum_{G} S|G\rangle(\pm)\langle G|S^{\dagger}$$

$$= \sum_{P} S|P\rangle\langle P|S^{\dagger} \; . \qquad \text{(I57)}$$

The ghost part can be shown to vanish when considered as an operator on the physical Hilbert space.

I showed you how such a proof can be set up by showing equivalence with a "unitary gauge", containing no ghost space G. In practice a more accurate procedure requires intermediate gauges, having only ghost particles with very high masses. Our cutting relations enable us to understand unitarity in all channels where the energy does not exceed the values of these masses.

I hope to have shown in this lecture why there is a strong theoretical argument (quite independent of the impressive experimental indications) in favor of a gauge theoretical structure of any field theory containing particles with spin one.

REFERENCES

More details on the subject of this lecture can be found in:

G. 't Hooft and M. Veltman, "DIAGRAMMAR", CERN report 73-9.