THE CONCEPTUAL BASIS OF QUANTUM FIELD THEORY

Gerard 't Hooft

Institute for Theoretical Physics
Utrecht University, Leuvenlaan 4
3584 CC Utrecht, the Netherlands

and

Spinoza Institute
Postbox 80.195
3508 TD Utrecht, the Netherlands

e-mail: `g.thooft@phys.uu.nl`
internet: `http://www.phys.uu.nl/~thooft/`

Version 28/01/05

January 28, 2005

Abstract

Relativistic Quantum Field Theory is a mathematical scheme to describe the sub-atomic particles and forces. The basic starting point is that the axioms of Special Relativity on the one hand and those of Quantum Mechanics on the other, should be combined into one theory. The fundamental ingredients for this construction are reviewed. A remarkable feature is that the construction is not perfect; it will not allow us to compute all amplitudes with unlimited precision. Yet in practice this theory is more than accurate enough to cover the entire domain between the atomic scale and the Planck scale, some 20 orders of magnitude.

Contents

1.   Introduction to the notion of quantized fields

Quantum Field Theory is one of those cherished scientific achievements that have become considerably more successful than they should have, if one takes into consideration the apparently shaky logic on which it is based. With awesome accuracy, all known subatomic particles appear to obey the rules of one example of a quantum field theory that goes under the uninspiring name of "The Standard Model". The creators of this model had hardly anticipated such a success, and one can rightfully ask to what it can be attributed.

We have long been aware of the fact that, in spite of its successes, the Standard Model cannot be *exactly* right. Most quantum field theories are not asymptotically free, which means that they cannot be extended to arbitrarily small distance scales. We could easily cure the Standard Model, but this would not improve our understanding at all, because we know that, at those extremely tiny distance scales where the problems would become relevant, a force appears that we cannot yet describe unambiguously: the gravitational force. It would have to be understood first.

Perhaps this is the real strength of Quantum Field Theory: we know where its limits are, and these limits are far away. The gravitational force acting between two subatomic particles is tremendously weak. As long as we disregard *that*, the theory is perfect. And, as I will explain, its internal logic is not shaky at all.

Subatomic particles all live in the domain of physics where spins and actions are comparable to Planck's constant $\hbar$. One obviously needs Quantum Mechanics to describe them. Since the energies available in sub-atomic interactions are comparable to, and often larger than, the rest mass energy mc^2 of these particles, they often travel with velocities close to that of light, c, and so relativistic effects will also be important. Thus, in the first half of the twentieth century, the question was asked:

"How should one reconcile Quantum Mechanics with Einstein's theory of Special Relativity?"

As we shall explain, Quantum Field Theory is the answer to this question.

Our first intuitions would be, and indeed were, quite different[1]. One would set up abstract Hilbert spaces of states, each containing fixed or variable numbers of particles. Subsequently, one would postulate a consistent scheme of interactions. What would 'consistent' mean? In Quantum Mechanics, we have learned how to describe a process where we start with a certain number of particles that are all far apart but moving towards one another. This is the 'in' state $|\psi\rangle_{\text{in}}$. After the interaction has taken place, we end up with particles all moving away from one another, a state $|\psi'\rangle_{\text{out}}$. The probability that a certain in-state evolves into a given out-state is described by a quantum mechanical transition amplitude, $_{\text{out}}\langle\psi'|\psi\rangle_{\text{in}}$. The set of all such amplitudes in the vector spaces formed by all in- and out-states is called the *scattering matrix*. One can ask how to construct the scattering matrix in such a way that (i) it is invariant under Lorentz transformations, and (ii) obeys the strict laws of *quantum causality*. By 'quantum causality' we mean that under no circumstance a measurable effect may proceed with a velocity faster than that of light. In practice, this means that one must demand that any set of local operators $\mathcal{O}_i(\mathbf{x}, t)$ obeys commutation rules such that the commutators $[\mathcal{O}_i(\mathbf{x}, t), \mathcal{O}_j(\mathbf{x}', t')]$ vanish as soon as the vector $(\mathbf{x} - \mathbf{x}', t - t')$ is space-like. One can show that this implies that the scattering matrix must obey *dispersion relations*.

This is indeed how physicists started to think about their problem. But how should one construct such a scattering matrix? Does any systematic procedure exist?

A quantized *field* may seem to be something altogether different, yet it does appear to allow for the construction of an interacting medium that does obey the laws of Lorentz invariance and causality. The local operators can be constructed from the fields. All we then have to do is to set up schemes of relativistically covariant field equations, such as Maxwell's laws. Even the introduction of non-linear terms in these equations appears to be straightforward, and if we were to subject such systems to a mathematically well-defined procedure called "quantization", we would have candidates for a solution to the aforementioned problem.

Realizing that the energy in a quantized field comes in quantized energy packages, which in all respects behave like elementary particles, and, conversely, realizing that operators in the form of fields could be defined also when one starts up with Hilbert spaces consisting of elementary particles, it was discovered that quantized fields do indeed describe subatomic particles. Subsequently, it was discovered that, in a quantized field, the number of ways in which interactions can be introduced (basically by adding non-linear terms in the field equations), is quite limited. Quantization requires that all interactions can be given in the form of a Lagrange function $\mathcal{L}$; relativity requires this $\mathcal{L}$ to be Lorentz-invariant, and, most strikingly, self-consistency of Quantum Field Theory then provides

further restrictions, which leads to the possibility of writing down a complete list of all possible interactions. The Standard Model is just one element of this list.

The scope of this concise treatise on Quantum Field Theory is too limited to admit detailed descriptions of all technical details. Instead, special emphasis is put on the conceptual issues that arise when addressing the numerous questions and problems associated with this doctrine. One *could* use this text to *learn* Quantum Field Theory, but for many technical details, more literature must be consulted.[2] We also limited ourselves to applications of Quantum Field Theory in elementary particle physics. There are many examples in low-temperature physics where these and similar techniques are useful, but they will not be addressed here.

2. Scalar fields

2.1. Classical Theory. Feynman rules

A field is here taken to mean a physical variable that is a function of space-time coordinates $x = (\mathbf{x}, t)$. In order for our theories to be in accordance with special relativity, we will have to specify how a field transforms under a homogeneous Lorentz transformation,

$$x' = Lx \ . \tag{2.1}$$

If a field ϕ transforms as

$$\phi'(x) = \phi(x') \ , \tag{2.2}$$

then ϕ is called a scalar field. The *improper* Lorentz transformations, such as parity reflection P and time reversal T, are of lesser importance since we know that Nature is not exactly invariant under those.

Let us first restrict ourselves to *real* scalar fields; generalization to the case where fields are denoted by complex numbers will be straightforward. Upon quantization, scalar fields will come in energy packets that behave as spinless Bose-Einstein particles, such as π^0, $\pi^\pm$ and η^0. Conceptually, the scalar field is the easiest to work with, but in section 9 we shall find reasons why other kinds of fields can actually improve the internal consistency of our theories.

Lorentz-invariant field equations typically take the form[1]

$$(\partial_\mu^2 - m_{(i)}^2)\phi_i = F_i(\phi) \ ; \qquad \partial_\mu^2 \equiv \vec{\partial}_x^2 - \partial_t^2 \ . \tag{2.3}$$

---

[1]We use summation convention: repeated indices that are not put between brackets are automatically summed over. Greek indices μ are Lorentz indices taking 4 values, Latin indices $i, j, \cdots$ run from 1 to 3. Our metric convention is $g_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$.

Here, the index i labels different possible species of scalar fields, and $F_i(\phi)$ could be any function of the field(s) $\phi_j(x)$. Usually, however, we assume that there is a potential function $V^{\text{int}}(\phi)$, such that $F_i(\phi)$ is the gradient of V^{int}, and furthermore we assume that V^{int} is a polynomial whose degree is at most four:

$$V^{\text{int}}(\phi) \;\; = \;\; \tfrac{1}{6}g_{ijk}\phi_i\phi_j\phi_k + \tfrac{1}{24}\lambda_{ijk\ell}\,\phi_i\phi_j\phi_k\phi_\ell \;;$$

$$F_i(\phi) \;\; = \;\; \frac{\partial V^{\text{int}}(\phi)}{\partial \phi_i} \;\; = \;\; \tfrac{1}{2}g_{ijk}\phi_j\phi_k + \tfrac{1}{6}\lambda_{ijk\ell}\,\phi_j\phi_k\phi_\ell \;, \tag{2.4}$$

where g and λ must be totally symmetric under all permutations of their indices.[1] This is actually a limitation on the forms that $F_i(\phi)$ can take. Without this limitation, we would not have a conserved energy, and quantization of the theory would not be possible. Later, we will see why higher terms in the polynomial are not permitted (section 7).

In order to understand the general structure of the classical solutions to this set of equations, we temporarily add a function $-J_i(x)$ to $F_i(\phi)$ in Eq. (2.3). Subsequently, we expand the solution in powers of $J_i(x)$:

$$(m_{(i)}^2 - \partial_\mu^2)\phi_i(x) = J_i(x) - \frac{\partial}{\partial \phi_i}V^{\text{int}}(\phi(x)) \;;$$

$$\phi_i(x) = \phi_i^{(1)}(x) + \phi_i^{(2)}(x) + \psi_i^{(3)}(x) + \cdots$$

$$= \int \mathrm{d}^4 y\, G_{ij}(x-y)\Big(J_j(y) - F_j\big(\phi^{(1)}(y) + \phi^{(2)}(y) + \phi^{(3)}(y) + \cdots\big)\Big). \tag{2.5}$$

The function $G_{ij}(x-y)$ is a solution to the equation

$$(m_i^2 - \partial_\mu^2)G_{ij}(x-y) = \delta_{ij}\delta(x-y) \;, \tag{2.6}$$

while $\phi_i^{(2)}(x)$ is quadratic in $J_j(y)$ and $\phi^{(3)}(x)$ is cubic, and so on. Assembling terms of equal order in $J_j(y)$ we find a recursive procedure to solve the field equations (2.5). At the end of our calculation, we might set $J_i(x)$ equal to zero, or better, have J non-vanishing only in the far-away region where the particles originated, so that the J interaction is a simplified model for the machine that produced the particles in the far past. Indeed, in the quantum theory it will also turn out to be convenient to use J as a model for the particle detector at the end of the experiment.

We see that the solution to Eq. (2.5) can be written as the sum of a large number of terms. Each of these terms can be written in the form of a diagram, called a Feynman diagram. In these diagrams, we represent a space-time point as a dot, and the function $G_{ij}(x-y)$ as a line connecting x with y. The index i may be indicated at each line. A dot may either be associated with a term $J_i(y)$, or it is a three-point vertex associated with a coefficient g_{ijk} or a four-point vertex, going with a coefficient $\lambda_{ijk\ell}$. A typical Feynman diagram is sketched in Fig. 1.
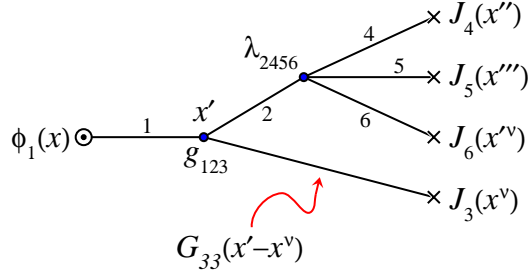
Figure 1: Example of a Feynman diagram for classical scalar fields

Observe the general structure of these diagrams. There are factors $\frac{1}{2}$, $\frac{1}{6}$, etc., which can easily be read off from the symmetries of the diagram. By construction, there are no closed loops: the diagram is simply connected. This will be different in the quantized theory.

There is one important issue to be addressed: the Green function $G_{ij}(x - y)$ is not completely determined by the equation (2.6): one may add arbitrary combinations of the solutions of the homogeneous equation $(m_i^2 - \partial_\mu^2)G_{ij}(x - y) = 0$. In Fourier space, this ambiguity is reflected in the fact that one has some freedom in choosing the integration curve C in the solution[2]

$$G_{ij}(x - y) = (2\pi)^{-4} \int_C \mathrm{d}^4k \, e^{ik\cdot(x-y)} \frac{\delta_{ij}}{k^2 + m_i^2} \; . \tag{2.7}$$

Our choice can be indicated by shifting the pole by an infinitesimal imaginary number, after which we choose the contour C to be along the real axis of all integrands. A reasonable choice is

$$G_{ij}^+(x - y) = (2\pi)^{-4} \int \mathrm{d}^4k \, e^{ik\cdot(x-y)} \frac{\delta_{ij}}{\mathbf{k}^2 - (k^0 + i\varepsilon)^2 + m_{(i)}^2} \; , \tag{2.8}$$

where ε is an infinitesimal, positive number. With this choice, the integration contour in the complex k^0 plane can be shifted such that the imaginary part of k^0 can be given an arbitrarily large positive value, and from this one deduces that the Green function will vanish as soon as the time difference, $x^0 - y^0$, is negative. This Green function, called the *forward* Green function, gives our expressions the desired causality structure: There are obviously no effects that propagate backwards in time, or indeed faster than light.

The converse choice, $G^-(x-y)$, gives us the backward solution. However, in the quantized theory, we will often be interested in yet another choice, the *Feynman propagator*,

---

[2]An inner product $k \cdot x$ stands for $\vec{k} \cdot \vec{x} - k^0 x^0$.

defined as

$$G_{ij}^F(x - y) = (2\pi)^{-4} \int \mathrm{d}^4 k \, e^{ik \cdot (x-y)} \frac{\delta_{ij}}{\mathbf{k}^2 - k^{0^2} + m_i^2 - i\varepsilon} \,, \tag{2.9}$$

where, again, the infinitesimal number $\varepsilon > 0$.

The rules to obtain the complete expansion of the solution can now be summarized as follows:

1) Each term can be depicted as a diagram consisting of points (vertices) connected by lines (called propagators). One end-point, ⊙———, corresponds to a point x where we want to know the field ϕ; the other end points, ———×, refer to factors $J(y^{(i)})$ for the corresponding points $y^{(i)}$, see Fig. 1.

2) There are no "closed loops". *i.e.* the diagrams must be simply connected (this will be different in the quantum theory).

3) There are vertices with three prongs (3-vertices), $\underset{i}{\overset{j}{\diagdown}}_{k}$, each being associated with a factor g_{ijk}, and vertices with four prongs (4-vertices), $\underset{i}{\overset{j}{\diagdown}}\!\!\!\!k_l$, each giving a factor $\lambda_{ijk\ell}$.

4) Each line connecting two points $x^{(1)}$ and $x^{(2)}$, $x^{(1)} \overset{i \quad k \quad j}{\bullet\!\!-\!\!-\!\!\bullet} x^{(2)}$, is associated with a factor $G_{ij}(x^{(1)} - x^{(2)})$ when we work in ordinary space-time (configuration space), or a factor

$$\frac{\delta_{ij}}{k^2 + m_i^2 - i\varepsilon} \,, \tag{2.10}$$

in momentum space (the reason for this $i\varepsilon$ choice will only become apparent in the quantized theory).

5) If we work in configuration space, we must integrate over all x values at each vertex except the one where ϕ was defined; if we work in momentum space, we must integrate over the k values, subject to the restriction of momentum conservation at each vertex: $k_{\text{out}} = \sum_{\text{in}} k_{\text{in}}$.

6) A 'combinatorial factor'. For the classical theories it is $1/N$, where N is the number of permutations of the source vertices that leave the diagram unaltered.

It is not difficult to generalize the rules for the case of higher polynomials in the interactions, but this will not be needed for the time being.

2.2. Spontaneous symmetry breaking. Goldstone modes

In the classical theory, the Hamilton density is

$$\mathcal{H}(\mathbf{x}, t) = \tfrac{1}{2}\dot{\phi}_i^2 + \tfrac{1}{2}(\vec{\partial}\phi_i)^2 + V(\phi) \ ; \qquad V(\phi) = \tfrac{1}{2}m_i^2\phi_i^2 + V^{\,\text{int}}(\phi) \ . \qquad (2.11)$$

The theory is invariant under the group of transformations

$$\phi_i'(x) = A_{ij}\phi_j(x) \ , \qquad (2.12)$$

if A is orthogonal and the potential function $V(\phi)$ is invariant under that group. The simplest example is the transformation $\phi \leftrightarrow -\phi$:

$$A = \pm 1 \ ; \qquad V = V(\phi^2) = \tfrac{1}{2}a\,\phi^2 + \frac{\lambda}{24}\phi^4 \ . \qquad (2.13)$$

There are two cases to consider:

i) $a > 0$. In this case, $\phi = 0$ is the absolute minimum of V. We write

$$a = m^2 \ , \qquad (2.14)$$

and find that m indeed describes the mass of the particle. All Feynman diagrams have an even number of external lines. Since, in the quantum theory, these lines will be associated with particles, we find that states with an odd number of particles can never evolve into states with an even number of particles, and *vice versa*. If we define the quantum number $P_C = (-1)^N$, where N is the number of ϕ particles, then we find that P_C is conserved during interactions.

ii) $a < 0$. In this case, we see that:

— trying to identify the mass of the particle using Eq. (2.14) yields the strange result that the mass would be purely imaginary. Such objects ("tachyons") are not known to exist and probably difficult to reconcile with causality, and furthermore:

— the configuration $\phi = 0$ does *not* correspond to the lowest energy configuration of the system. The lowest energy is achieved when

$$\phi = \pm F \ ; \qquad F^2 = -6a/\lambda \ . \qquad (2.15)$$

It is now convenient to rewrite the potential V as

$$V = \frac{\lambda}{24}(\phi^2 - F^2)^2 - C \ , \qquad (2.16)$$

where we did not bother to write down the value of the constant C, since it does not occur in the evolution equations (2.5). There are now *two equivalent vacuum states*, the minima of V. Choosing one of them, we introduce a new field variable $\tilde{\phi}$ to write

$$\phi \;\equiv\; F + \tilde{\phi} \;;$$
$$V \;=\; \frac{\lambda}{24}\tilde{\phi}^2(2F + \tilde{\phi})^2 \;=\; \frac{\lambda F^2}{6}\tilde{\phi}^2 + \frac{\lambda F}{6}\tilde{\phi}^3 + \frac{\lambda}{24}\tilde{\phi}^4 \;, \qquad (2.17)$$

and we see that

$a)$ for the new field $\tilde{\phi}$, the mass-squared $\tilde{m}^2 = \lambda F^2/3$ is positive, and

$b)$ a three-prong vertex appeared, with associated factor λF. The quantum number P_C is no longer apparently conserved.

This phenomenon is called 'spontaneous symmetry breaking', and it plays an important role in Quantum Field Theory.

Next, let us consider the case of a continuous symmetry. The prototype example is the $U(1)$ symmetry of a complex field. The symmetry group consists of the transformations $A(\theta)$, where θ is an angle:

$$\Phi \equiv \tfrac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \;; \qquad \Phi' = A(\theta)\Phi = e^{i\theta}\Phi \;, \qquad (2.18)$$

Again, the most general potential[3] invariant under these transformations is

$$V(\Phi, \Phi^*) = a\,\Phi^*\Phi + \tfrac{1}{2}\lambda(\Phi^*\Phi)^2 - C \;, \qquad (2.19)$$

In the case where the $U(1)$ symmetry is apparent, one can rewrite the Feynman rules to apply directly to the complex field Φ, noticing that one can write the potential V as a real function of the two independent variables Φ and Φ^*. With

$$\partial_\mu^2 \Phi = \frac{\partial V(\Phi, \Phi^*)}{\partial \Phi^*} \;, \qquad (2.20)$$

one notices that the Feynman propagators can be written with an arrow in them: an arrow points towards a point x where the function $\Phi(x)$ is called for, and away from a point x' where a factor $\Phi^*(x')$ is extracted from the potential V. At every vertex, as many arrows enter as they leave, and so, during an interaction, the total number of lines pointing forward in time minus the number of lines pointing backward is conserved. This

---

[3]Observe how we adjusted the combinatorial factors. The choices made here are the most natural ones to keep these coefficients as predictable as possible in future calculations.

is an additively conserved quantum number, to be interpreted as a 'charge' Q. According to *Noether's theorem*, every symmetry is associated to such a conservation law.

However, if $a < 0$, this $U(1)$ symmetry is spontaneously broken. Then we write

$$V = \tfrac{1}{2}\lambda(\Phi^*\Phi - F^2)^2 - C \ , \qquad F^2 \equiv -a/\lambda \ . \tag{2.21}$$

This time, the stable vacuum states form a closed circle in the complex plane of Φ values. Let us write

$$\begin{aligned}
\Phi &\equiv& F + \tilde{\Phi} \ ; \qquad \tilde{\Phi} \equiv \tfrac{1}{\sqrt{2}}(\tilde{\phi}_1 + i\tilde{\phi}_2) \ ; \\
V &=& \tfrac{1}{2}\lambda\Big(F(\tilde{\Phi} + \tilde{\Phi}^*) + \Phi^*\Phi\Big)^2 \\
&=& \lambda F\,\tilde{\phi}_1^2 + \frac{\lambda F}{\sqrt{2}}\tilde{\phi}_1(\tilde{\phi}_1^2 + \tilde{\phi}_2^2) + \frac{\lambda}{8}(\tilde{\phi}_1^2 + \tilde{\phi}_2^2)^2 \ .
\end{aligned} \tag{2.22}$$

The striking thing about this potential is that the mass term for the field $\tilde{\phi}_2$ is missing. The mass squared for the $\tilde{\phi}_1$ field is $\tilde{m}_1^2 = 2\lambda F$. The fact that one of the effective fields is massless is a fundamental consequence of the fact that we have spontaneous breakdown of a continuous symmetry. Quite generally, there is a theorem, called the *Goldstone theorem*:

If a continuous symmetry whose symmetry group has N independent generators, is broken down spontaneously into a (residual) symmetry whose group has N_1 independent generators, then $N - N_1$ massless effective fields emerge.

The propagators for massless fields obey Eq. (2.6) without the m^2 term, which gives these expressions an 'infinite range': such a Green's function drops off only slowly for large spatial or timelike separations. These massless oscillating modes are called 'Goldstone modes'.

2.3. Quantization of a classical theory

How does one "quantize" a field theory? In the old days of Quantum Mechanics, it was taught that "you take the Poisson brackets of the classical system, and replace these by commutators." Here and there, one had to readjust the *order* in which classical expressions emerge, when they are replaced by operators, but the rules appeared to leave no essential ambiguities. Indeed, *if* such a procedure is possible, one may get a quantum theory which reproduces the original classical system in the limit of vanishing $\hbar$. Also, the group of symmetry transformations under which the classical system was invariant, often re-emerges in the quantum system.

A *field theory*, however, has a strictly infinite set of physical degrees of freedom (the field values at every point in 3-space, or, the complete set of Fourier modes). More often than not, upon "quantization", this leads to infinities that render the theory ill-defined.

One has to formulate the notion of "quantization" much more carefully, going through several intermediate steps. Since, today, the answers to our questions are so well known, it is often forgotten how these answers can be derived rigorously and why they take the form they have. What is the strictly logical sequence of arguments?

First of all, it is unreasonable to expect that every classical field theory should have a quantum mechanical counterpart. What we wish to do, is construct *some* quantum system, its Hilbert space and its Hamiltonian, such that in one or more special limits, it reproduces a known classical theory. We demand certain properties of the theory, such as Lorentz invariance and causality, but most of all we demand that it be internally logically impeccable, allowing us to calculate how in such a system particles interact, under all imaginable circumstances. We will, however, continue to use the phrase 'quantization', meaning that we attempt to construct a quantum theory with a given classical field theory as its $\hbar \to 0$ limit.

Often, authors forget to mention the first, very important, step in this logical procedure: replace the classical field theory one wishes to quantize by a strictly finite theory. Assuming that physical structures smaller than a certain size will not be important for our considerations, we replace the continuum of three-dimensional space by a discrete but dense lattice of points. In the differential equations, we replace all derivatives $\partial/\partial x^i$ by finite ratios of differences: $\Delta/\Delta x^i$, where $\Delta\phi$ stands for $\phi(x + \Delta x) - \phi(x)$. In Fourier space, this means that wave numbers $\vec{k}$ are limited to a finite range (the Brillouin zone), so that integrations over $\vec{k}$ can never diverge.

If this lattice is sufficiently dense, the solutions we are interested in will hardly depend on the details of this lattice, and so, the classical system will resume Lorentz invariance and the speed of light will be the practical limit for the velocity of perturbances. If necessary, we can also impose periodic boundary conditions in 3-space, and in that case our system is completely finite. Finite systems of this sort allow for 'quantization' in the old-fashioned sense: replace the Poisson brackets by commutators. Note that we did not (yet) discretize time, so the Hamiltonian of our theory has the form

$$H = T + V = \sum_{x^a} \prod_{a=1}^{3} (\Delta x^a) \left(\tfrac{1}{2} \sum_i (\partial\phi_i/\partial t)^2 + \tfrac{1}{2} \sum_{i,a} \left(\frac{\Delta\phi_i}{\Delta x^a} \right)^2 + V(\phi) \right) . \qquad (2.23)$$

The canonical momenta associated to the fields $\phi_i(x)$ are

$$p^i(x) = (\partial\phi_i/\partial t) \prod_{a=1}^{3} (\Delta x^a) , \qquad (2.24)$$

and so, we will assume these to be operators obeying:

$$[\phi_i(x),\, \phi_j(x')] = 0 \; ; \qquad [p^i(x),\, p^j(x')] = 0 \; ; \qquad [\phi_i(x),\, p^j(x')] = i\delta_i^j\, \delta_{x,x'} . \qquad (2.25)$$

12

Now, we have to wait and see what happens in the limit of an infinitely dense space-lattice. Will, like the classical theory, our quantum concoction turn out to be Lorentz-invariant? How do we perform Lorentz transformations on physical states? This question turns out to be far from trivial to answer, but the answer is known. We first need some useful technical tools.

2.4. The Feynman path integral

The Feynman path integral is often introduced as an "infinite dimensional" integral. Again, we insist on at first keeping everything finite. Label the generalized coordinates (here the ϕ_i fields) as q_i. The momenta are p_i. The Hamiltonian (2.23) is of the conventional type (the volume elements $\prod_{a=1}^{3}(\Delta x^a)$ act as masses). For future use, we need a slightly more general one, a Hamiltonian that also contains pieces linear in the momenta p_i:

$$H = T + V \; ; \qquad T = \sum_i \frac{\left(p_i - A_i(\mathbf{q})\right)^2}{2\,m_{(i)}} \; ; \qquad V = V(\mathbf{q}) \; . \tag{2.26}$$

In principle, we keep the number n of coordinates and momenta finite, in which case there is no doubt that the differential equations in question have unique, finite solutions (assuming the functions A_i and V to be sufficiently smooth; indeed we will mostly work with polynomials). Consider the configuration states $|\mathbf{q}\rangle$ and the momentum states $|\mathbf{p}\rangle$. We have

$$\langle \mathbf{q}|\mathbf{q}'\rangle = \delta^n(\mathbf{q} - \mathbf{q}') \; , \qquad \langle \mathbf{p}|\mathbf{p}'\rangle = \delta^n(\mathbf{p} - \mathbf{p}') \; ; \qquad \langle \mathbf{q}|\mathbf{p}\rangle = (2\pi)^{-n/2} e^{ip_i q_i} \; . \tag{2.27}$$

Taking the order of the operators into account, we write for the kinetic energy

$$T = \sum_i \frac{p_i^2 - 2A_i p_i + A_i^2}{2\,m_{(i)}} + iW(\mathbf{q}) \; ;$$
$$W(\mathbf{q}) = \sum_i \frac{[A_i(\mathbf{q}),\,p_i]}{2i\,m_{(i)}} = \sum_i \frac{\partial_i A_i(\mathbf{q})}{2\,m_{(i)}} \; . \tag{2.28}$$

This enables us to compute swiftly the matrix elements

$$\langle \mathbf{q}|H|\mathbf{p}\rangle = \langle \mathbf{q}|\mathbf{p}\rangle(h(\mathbf{q}, \mathbf{p}) + iW(\mathbf{q})) \; ; \tag{2.29}$$
$$\langle \mathbf{p}|H|\mathbf{q}\rangle = \langle \mathbf{p}|\mathbf{q}\rangle(h(\mathbf{q}, \mathbf{p}) - iW(\mathbf{q})) \; , \tag{2.30}$$

where $h(\mathbf{q}, \mathbf{p})$ is the *classical* Hamiltonian as a function of the two sets of variables $\mathbf{q}$ and $\mathbf{p}$.

13

The *evolution operator* $U(t, \delta t)$ for a short time interval δt is

$$U(t, \delta t) \quad = \quad e^{-iH(t)\delta t} \quad = \quad \mathbb{I} - iH\,\delta t + \mathcal{O}(\delta t)^2 \ . \tag{2.31}$$

Its matrix elements between states $\langle \mathbf{p} |$ and $| \mathbf{q} \rangle$ are easy to derive now:

$$
\begin{aligned}
\langle \mathbf{p} | U(t, \delta t) | \mathbf{q} \rangle \quad &= \quad \langle \mathbf{p} | \mathbf{q} \rangle - i\delta t \langle \mathbf{p} | H | \mathbf{q} \rangle + \mathcal{O}(\delta t)^2 \\
&= \quad (2\pi)^{-n/2} e^{-ip_i q_i} \Big(1 - i\delta t \{ h(\mathbf{q}, \mathbf{p}) - iW(\mathbf{q}) \} + \mathcal{O}(\delta t)^2 \Big) \\
&= \quad (2\pi)^{-n/2} \exp\Big(-i\mathbf{p} \cdot \mathbf{q} - i\delta t \{ h(\mathbf{q}, \mathbf{p}) - iW(\mathbf{q}) \} + \mathcal{O}(\delta t)^2 \Big) \ .
\end{aligned}
\tag{2.32}
$$

What makes this expression very useful is the fact that it does not become singular in the limit $\delta t \downarrow 0$. The momentum-momentum and the coordinate-coordinate matrix elements do become singular in that limit.

Next, let us consider a finite time interval T. The evolution operator over that time interval can formally be viewed as a sequence of many evolution operators over short time intervals δt, with $T = N\,\delta t$. Using closure, both in $\mathbf{p}$ space and in $\mathbf{q}$ space, at all time intervals,

$$\mathbb{I} \quad = \quad \int \mathrm{d}^n \mathbf{q}\, | \mathbf{q} \rangle \langle \mathbf{q} | \quad = \quad \int \mathrm{d}^n \mathbf{p}\, | \mathbf{p} \rangle \langle \mathbf{p} | \ , \tag{2.33}$$

we can write

$$
\begin{aligned}
|\psi(\mathbf{q}_N, T)\rangle \quad &= \quad \langle \mathbf{q}_N | U(0, T) | \psi(0) \rangle \quad = \quad \int \mathrm{d}^n \mathbf{q}_0 \int \mathrm{d}^n \mathbf{p}_0 \cdots \int \mathrm{d}^n \mathbf{q}_{N-1} \int \mathrm{d}^n \mathbf{p}_{N-1} \\
&\quad \langle \mathbf{q}_N | \mathbf{p}_{N-1} \rangle \, \langle \mathbf{p}_{N-1} | U(t_{N-1}, \delta t) | \mathbf{q}_{N-1} \rangle \, \langle \mathbf{q}_{N-1} | \mathbf{p}_{N-2} \rangle \\
&\quad \langle \mathbf{p}_{N-2} | U(t_{N-2}, \delta t) | \mathbf{q}_{N-2} \rangle \, \cdots \, \langle \mathbf{p}_0 | U(0, \delta t) | \mathbf{q}_0 \rangle \, \langle \mathbf{q}_0 | \psi(0) \rangle \ .
\end{aligned}
\tag{2.34}
$$

Plugging in Eq. (2.32), we see that

$$
\begin{aligned}
|\psi(\mathbf{q}_N, T)\rangle \quad &= \quad \Big(\prod_{\tau=0}^{N-1} \int \mathrm{d}^n \mathbf{q}_\tau \int \mathrm{d}^n \mathbf{p}_\tau \frac{e^{-W(\mathbf{q}_\tau)\delta t}}{(2\pi)^n} \Big) \times \\
&\quad \exp i \sum_{\tau=0}^{N-1} \delta t \Big(\mathbf{p}_\tau \frac{\mathbf{q}_{\tau+1} - \mathbf{q}_\tau}{\delta t} - h(\mathbf{q}_\tau, \mathbf{p}_\tau, t_\tau) \Big) \, \langle \mathbf{q}_0 | \psi(0) \rangle \ .
\end{aligned}
\tag{2.35}
$$

Define

$$\dot{\mathbf{q}}_\tau \equiv \frac{\mathbf{q}_{\tau+1} - \mathbf{q}_\tau}{\delta t} \ , \tag{2.36}$$

and

$$L(\mathbf{p}, \mathbf{q}, \dot{\mathbf{q}}, t) = \mathbf{p} \cdot \dot{\mathbf{q}} - h(\mathbf{q}, \mathbf{p}, t) \ , \tag{2.37}$$

14

and the *measure*

$$\prod_{\tau=0}^{N-1} \int \mathrm{d}^n\mathbf{q}_\tau \int \mathrm{d}^n\mathbf{p}_\tau \frac{e^{-W(\mathbf{q}_\tau)\delta t}}{(2\pi)^n} \equiv \int \mathcal{D}\mathbf{q}\mathcal{D}\mathbf{p} \ , \tag{2.38}$$

then we obtain an expression that seems to be easy to extend to infinitely fine grids in the time variable:

$$\langle \mathbf{q}_N|\psi(T)\rangle = \int \mathcal{D}\mathbf{q}\mathcal{D}\mathbf{p} \left(\exp i \sum_{\tau=0}^{N-1} \delta t \, L(\mathbf{p},\, \mathbf{q},\, \dot{\mathbf{q}},\, t) \right) \langle \mathbf{q}_0|\psi(0)\rangle \ . \tag{2.39}$$

In these expressions, we actually allowed the parameters in the Hamiltonian H and the Lagrangian L to depend explicitly on time t, so as to expose the physical structure of these expressions. Note that

$$L(\mathbf{p},\, \mathbf{q},\, \dot{\mathbf{q}},\, t) = -\sum_i \frac{(p_i - A_i - m_{(i)}\dot{q}_i)^2}{2\, m_{(i)}} - V(\mathbf{q}) + \sum_i (A_i \dot{q}_i + \tfrac{1}{2}m_{(i)}\dot{q}_i^2) \ , \tag{2.40}$$

and the integrals over all momentum variables are easy to perform, giving some constant that only depends on the masses $m_{(i)}$:

$$\langle \mathbf{q}_N|\psi(T)\rangle = \int \mathcal{D}\mathbf{q} \exp \left(i \sum_{\tau=0}^{M-1} \delta t \, L(\mathbf{q},\dot{\mathbf{q}},t) \right) \langle \mathbf{q}_0|\psi(0)\rangle \ , \tag{2.41}$$

with

$$L(\mathbf{q},\dot{\mathbf{q}},t) = T - V \ ; \qquad T = \sum_i (\tfrac{1}{2}m_{(i)}\dot{q}_i^2 + A_i \dot{q}_i) \ ;$$

$$\mathcal{D}\mathbf{q} = e^{-\sum_\tau W(\mathbf{q}_\tau)\,\delta t} \prod_{\tau=0}^{N-1} \left(\mathrm{d}^n\mathbf{q}_\tau \prod_i \left(\frac{m_{(i)}}{2\pi\,\delta t} \right)^{\frac{1}{2}} \right) \ . \tag{2.42}$$

Actually, $L(\mathbf{q},\dot{\mathbf{q}},t)$ is obtained from $L(\mathbf{p},\mathbf{q},\dot{\mathbf{q}},t)$ by *extremizing* the latter with respect to $\mathbf{p}$:

$$\frac{\partial}{\partial p_i}L(\mathbf{p},\mathbf{q},\dot{\mathbf{q}},t) = 0 \ ; \qquad \dot{q}_i = \frac{\partial h(\mathbf{q},\mathbf{p},t)}{\partial p_i} \ . \tag{2.43}$$

This is exactly the standard relation between Lagrangian and Hamiltonian of the classical theory. So, L is indeed the Lagrangian.

If the continuum limit exists, the exponent in Eq. (2.41) is exactly i times the classical action,

$$S = \int \mathrm{d}t L(\mathbf{q},\dot{\mathbf{q}},t) \ . \tag{2.44}$$

15

It is tempting to assume that the $\mathcal{O}(\delta t)^2$ terms in Eq. (2.31) disappear in the limit; after all, they are only multiplied by factors $N \approx C/\delta t$. In that case, the evolution operator in Eq. (2.41) clearly takes the form of *an integral over all paths* going from $\mathbf{q}_0$ to $\mathbf{q}_N$. This is Feynman's path integral. In the case of a field theory, one considers the field defined on a lattice in space, and since the path integral starts with a lattice in the time variable, we end up dealing with a lattice in space and time. In conclusion:

> The evolution operator in a field theory is described by first rephrasing the theory on a dense lattice in space-time. Replacing partial derivatives by the corresponding finite difference ratios, one writes an expression for the *action* S of the theory. Normally, it can be written as an integral over a Lagrange density, $\mathcal{L}(\phi, \partial_\mu \phi)$. The evolution operator of the theory is obtained by integrating e^{iS} over all field configurations $\phi(\mathbf{x}, t)$ in a given space-time patch. The integration measure is defined from Eq. (2.42).

The A_i terms, linear in the time derivatives, do not play a role in scalar field theories but they do in vector theories, and the fact that they occur in the measure (2.42) is usually ignored. Indeed, in most cases, $W(\mathbf{q})$ vanishes, but we must be aware that it might cause problems in some special cases. We ignore the W term for the time being.

2.5. Feynman rules for the quantized theory

The Feynman rules for quantized field theories were first derived by careful analysis of perturbation theory. Writing the quantum Hamiltonian H as $H = H_0 + H^{\text{int}}$, one assembles all terms bilinear in the fields and their derivatives in H_0 and performs the perturbation expansion for small values of H^{int}. This leads to a set of calculation rules very similar to the rules derived for a classical theory, see subsection 2.1. Most of these rules (but not everything) can now most elegantly be derived from the path integral.

Let us first derive these rules for computing a *finite dimensional* integral of the type (2.41). Although often our action will not contain terms linear in the variables $q_i(t)$, we do need such terms now, so, if necessary, we add them by hand, only to remove them at the end of the calculations. There is no need to indicate the time variable t explicitly; we absorb it in the indices i. the action is then

$$S(\mathbf{q}) = \sum_{\mathbf{x},t} \mathcal{L}(\mathbf{x}, t) = J_i q_i - \tfrac{1}{2} M_{ij} q_i q_j - \tfrac{1}{6} A_{ijk} q_i q_j q_k - \tfrac{1}{24} B_{ijk\ell} q_i q_j q_k q_\ell . \qquad (2.45)$$

To calculate $\int \mathrm{d}^N \mathbf{q}\, e^{iS(\mathbf{q})}$, we keep only the bilinear part (the term with the coefficients M_{ij}) inside the exponent, and expand the exponent of all other terms:

$$_{\text{out}}\langle 0|0\rangle_{\text{in}} = C \int \mathrm{d}^N \mathbf{q} \Big(\exp\big(-\tfrac{1}{2} i M_{ij} q_i q_j \big) \Big) \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} \frac{1}{k!\ell!m!} \times$$

16

$$(iJ_{i_1}q_{i_1})\cdots(iJ_{i_k}q_{i_k})\ (-\tfrac{i}{6}A_{i_1j_1k_1}\ q_{i_1}q_{j_1}q_{k_1})\cdots(-\tfrac{i}{6}A_{i_\ell j_\ell k_\ell}\ q_{i_\ell}q_{j_\ell}q_{k_\ell})$$
$$(-\tfrac{i}{24}B_{i_1j_1k_1\ell_1}\ q_{i_1}q_{j_1}q_{k_1}q_{\ell_1})\cdots(-\tfrac{i}{24}B_{i_mj_mk_m\ell_m}\ q_{i_m}q_{j_m}q_{k_m}q_{\ell_m})\ . \tag{2.46}$$

(C is a constant not depending on the coefficients, but only on their dimensionality).

We can calculate all of these integrals if we know how to do the J terms. These however can be done to all orders since we know exactly how to do the Gaussian integral

$$\int \mathrm{d}^N\mathbf{q}\exp i(-\tfrac{1}{2}M_{ij}q_iq_j + J_iq_i) = \frac{(2\pi)^{\frac{N}{2}}}{(\det(M))^{\frac{1}{2}}}\exp\left(\tfrac{1}{2}iJ_iM_{ij}^{-1}J_j\right) \ =$$
$$C\sum_{k=0}^{\infty}\frac{1}{k!}\left(\tfrac{1}{2}iJ_{i_1}M_{i_1j_1}^{-1}J_{j_1}\right)\cdots\left(\tfrac{1}{2}iJ_{i_k}M_{i_kj_k}^{-1}J_{j_k}\right)\ . \tag{2.47}$$

This expression tells us how to do the integrals in Eq. (2.46) by collecting terms that go with given powers of J_i. The outcome of this calculation can be summarized in a concise way:

1) Each term can be depicted as a diagram consisting of points (vertices) connected by lines (propagators). The lines may end at points i, $\longrightarrow\!\!\times$, which refer to factors J_i.

2) There are vertices with three prongs (3-vertices), $\underset{}{\overset{}{-i\!\!<^{j}_{k}}}$, each being associated with a factor A_{ijk}, and vertices with four prongs (4-vertices), $i\!\!<^{j}_{l}\!\!k$, each giving a factor $B_{ijk\ell}$.

3) Each line connecting two points i and j, is associated with a factor M_{ij}^{-1}.

4) In contrast with the classical theory, however, the diagrams may contain disconnected pieces, or multiply connected parts: closed loops. See Fig. 2.

5) There are combinatorial factors arising from the coefficients such as $k!$ in Eq. (2.46). One can gain experience in deriving these factors; they follow directly from the symmetry structure of a diagram. This technical detail will not be further addressed here.

Apparently nothing changes if one re-inserts the $(\mathbf{x}, t)$ dependence of these coefficients, when the variables q_i are replaced by the fields $\phi_i(\mathbf{x}, t)$, and the action by that of a field theory:

$$S = \int \mathrm{d}^4x\mathcal{L}(\mathbf{x}, t)\ ;$$
$$\mathcal{L}(\mathbf{x}, t) = -\tfrac{1}{2}(\partial_\mu\phi_i)^2 - \tfrac{1}{2}m_{(i)}^2\phi_i^2 - V(\phi) + J_i(x)\phi_i(x)\ . \tag{2.48}$$

17

The rules are as in Subsection 2.1, with the only real distinction that, in the quantum theory, diagrams with closed loops in them contribute. These diagrams may be regarded as the "quantum corrections" to the classical field theory. The disconnected diagrams mentioned under point (4), arise for technical reasons that we will not further elaborate; in practical calculations they may usually be ignored.
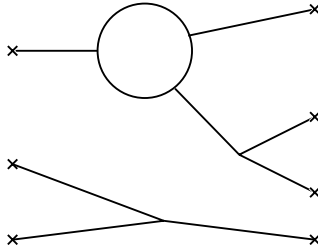

Figure 2: Example of a Feynman diagram for quantized scalar fields

At one point, however, we made an omission: the overall constant C was not computed. It comes from the cancellation of two coefficients (the one in the measure and the one coming from the Gaussian integrals) each of which tend to infinity in the limit of an infinitely dense grid. In most cases, we are not interested in this coefficient (it refers to vacuum-energy), but this does imply that more is needed to extract relevant physical information from these Feynman diagrams. Fortunately, this deficit is easy to cure. The "source insertions", $J_i(x)\phi_i(x)$ can serve as a model both for the production and for the detection of particles. Let both $|0\rangle_{\text{in}}$ and $|0\rangle_{\text{out}}$ be the vacuum, or ground state of the theory. At early times, the insertion $-J(\mathbf{x}, t)\phi$ in the Hamiltonian acts on this vacuum state to excite it into the initial state we are interested in. By differentiating with respect to J, we can reach any initial state we want to consider. Similarly, at the end of the experiment, at late times, $J\phi$ can link the particle state that we wish to detect to the final vacuum state. In short, differentiating with respect to $J(\mathbf{x}, t)$ gives us any matrix element that we wish to study. This is easier than one might think: J_i refers to particles of type i, and if we give it the same space-time dependence as the wave function of the particle we want to see (put it on the 'mass shell' of that particle), then we can be sure that there will be no contamination from unwanted particle states. One only has to check the normalization, but also that is not hard: we adjust the 1-particle to 1-particle amplitude to be one; a single particle cannot scatter (it could be unstable, but that is another matter). The constant C always drops out of these calculations.

An important point is the ambiguity of the inverse matrix M^{-1}. As in the classical case, there are homogeneous solutions, so, if we work in momentum space, there will be the question how to integrate around the poles of the propagator. The $i\varepsilon$ prescription mentioned in subsection 2.1 is now imperative. This is explained as follows. Consider the

propagator in position space, and choose its poles situated as follows:

$$\int \mathrm{d}^4k \frac{e^{i\mathbf{k}\cdot\mathbf{x}-ik^0t}}{m^2 + \mathbf{k}^2 - k^{0\,2} - i\varepsilon} \; ; \qquad \varepsilon \downarrow 0 \; . \tag{2.49}$$

The poles are at $k^0 = \pm(\sqrt{m^2 + \mathbf{k}^2} - i\varepsilon)$. Now consider this propagator at time $t = -T + i\beta$ with both T and β large. Since β is large, the choice of the contour at negative k^0 is immaterial, since the exponential there is very small. At positive k^0, we choose the contour to go above the pole, so the imaginary part of k^0 is chosen positive. We see that then the exponential vanishes rapidly at negative time. In short, our propagator tends to zero if the time t tends to $-T + i\beta$ when both T and β are large and positive. The same holds for $t \to +T - i\beta$. Indeed, we want our evolution operator to be dominated by the empty diagram in these two limits. Write:

$$\langle \psi | U(0, +T - i\beta) | \psi' \rangle \;=\; \sum_E \langle \psi | E \rangle \exp(-iET - \beta E) \langle E | \psi' \rangle \; , \tag{2.50}$$

where $|E\rangle$ are the energy eigenstates. At large β, the vacuum state should dominate. Conversely, if we consider evolution *backwards* in time, the other $i\varepsilon$ prescription is needed. One then works with the Feynman rules for the inverse, or the complex conjugate, of the scattering matrix.

Now, we are in a position to add the prescription how to identify the external lines (the lines sticking out of the diagram) with in- and out-going particles. For an ingoing particle, we use a source function $J(x)$ whose Fourier components emit a *positive* amount of energy k^0. For an out-going particle the source emits a *negative* k^0. According to the rules formulated above, these sources would be connected to the rest of the diagram by propagators, in Fourier space $(k_\mu^2 + m^2 - i\varepsilon)^{-1}$. Since the in- and out-going particles have $k_\mu^2 + m^2 = 0$, we must take the *residue of the pole*. In practice, this means that we have to remove the external propagators, a procedure called 'amputation'. One then still has to establish a normalization factor. This factor is most easily obtained by checking unitarity of the scattering matrix, using the optical theorem. At first sight, this seems to be just a simple numerical coefficient, but there is a slight complication at higher orders, when self-energy corrections affect the propagator. These corrections also remove *unstable particles* from the physical scattering matrix. We return to this in Section 6. The complete Feynman rules are listed in subsection 4.4

3. Spinor fields

3.1. The Dirac equation

The fields introduced in the previous section can only be used to describe particles with spin 0. In a quantum theory, particles can come in any representation of the *little group*,

which is the subgroup of the inhomogeneous Lorentz group that leaves the 4-momentum of a particle unaffected. For massive particles in ordinary space, this is the group of rotations of a three-vector, $SO(3)$. Its representations are labelled by either an integer ≥ 0, or an integer $+\frac{1}{2}$, representing the total spin of a particle. So, next in line are the particles with spin $\frac{1}{2}$. The wave function for such a particle has two components, one for spin up and one for spin down. Therefore, to describe a relativistic theory with such particles, we should use a two-component field obeying a relativistically covariant field equation. Paul Dirac was the first to find an appropriate relativistically covariant equation for a free particle with spin $\frac{1}{2}$:

$$(m + \sum_{\mu} \gamma^{\mu} \partial_{\mu}) \psi(x) = 0 , \tag{3.1}$$

but the field $\psi(\mathbf{x}, t)$ has *four complex* components. Here, γ^{μ}, $\mu = 0, 1, 2, 3$, are four 4×4 matrices, obeying

$$\{\gamma^{\mu}, \gamma^{\nu}\} = \gamma^{\mu} \gamma^{\nu} + \gamma^{\nu} \gamma^{\mu} = 2g^{\mu\nu} ; \qquad \gamma_{\mu}^{\dagger} = g_{\mu\nu} \gamma^{\nu} . \tag{3.2}$$

In contrast to the scalar case, the Dirac equation is *first order* in the space- and time-derivatives, and furthermore, one could impose a 'reality condition' (Majorana condition) on the fields, of the form

$$\psi(x) = C\psi^{*}(x) , \qquad \gamma^{\mu} C = C(\gamma^{\mu})^{*} , \quad \mu = 0, 1, 2, 3. \tag{3.3}$$

These two features combined give the Dirac field the same multiplicity as two scalar fields. Usually, we do not impose the Majorana condition, so that the Dirac field is truly complex, having a conserved $U(1)$ charge much like a complex scalar field.

We briefly recapitulate the most salient features of the Dirac equation. The 4×4 Dirac matrices can conveniently be expressed in terms of two commuting sets of Pauli matrices, σ_a and τ_a. Define

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{3.4}$$

and similarly for the τ matrices, except that they act in different spaces: a Dirac index is then viewed as a pair $(i\alpha)$ of indices i and α, such that the matrices σ_a act on the first index i, and the matrices τ_A act on the indices α. We have:

$$\sigma_a \sigma_b = \delta_{ab} + i\varepsilon_{abc} \sigma_c , \qquad \tau_A \tau_B = \delta_{AB} + i\varepsilon_{ABC} \tau_C , \qquad [\sigma_a, \tau_B] = 0 . \tag{3.5}$$

Define (with the convention $g_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$)

$$\gamma^1 = \sigma_1 \tau_1 , \quad \gamma^2 = \sigma_2 \tau_1 , \quad \gamma^3 = \sigma_3 \tau_1 , \quad \gamma^0 = -i\tau_3 . \tag{3.6}$$

20

The matrix C in Eq. (3.3) is then:

$$C = \gamma_2 \gamma_4 . \tag{3.7}$$

In the non-relativistic limit, the Dirac equation reads

$$(m + i\gamma^\mu k^\mu)\psi \;\approx\; (m - i\gamma^0 k^0)\psi \;\approx\; m(1 - \tau_3)\psi \;=\; 0 , \tag{3.8}$$

so that only two of the four field components survive (the ones with $\tau_3|\psi\rangle = |\psi\rangle$).

3.2.   Fermi-Dirac statistics

At this point, we could now attempt to pursue our fundamental quantization program: produce the Poisson brackets of the system, replace these by commutators, rewrite the Hamiltonian of the system in operator form, and solve the resulting Schrödinger equation.

Unfortunately, if one uses ordinary (commuting) numbers, this does not work. The Lagrangian associated to the Dirac equation will read

$$L = \int \mathrm{d}^3\vec{x}\, \mathcal{L}(x) ; \qquad \mathcal{L}(x) = -\overline{\psi}(x)(m + \sum_{\mu=0}^{4} \gamma^\mu \partial_\mu)\psi(x) , \tag{3.9}$$

and the canonical procedure would give as momentum fields:

$$p_\psi(\vec{x}) = \frac{\partial \mathcal{L}}{\partial(\partial_0 \psi(\vec{x}))} = \overline{\psi}(\vec{x})\gamma^0 , \qquad p_{\overline{\psi}}(\vec{x}) = 0 . \tag{3.10}$$

From this, one finds the Hamiltonian:

$$H = \int \mathrm{d}^3\vec{x}\, \mathcal{H}(\vec{x}) ; \qquad \mathcal{H}(\vec{x}) = p_\psi\, \dot{\psi} - \mathcal{L}(\vec{x}) = \overline{\psi}(x)(m + \sum_{i=1}^{3} \gamma^i \partial_i)\psi(x) , \tag{3.11}$$

Here, the index i is a spatial one, runnunig from 1 to 3. This, however, is not bounded from below! Such a quantum theory would not possess a vacuum state, and hence be unsuitable as a model for Nature.

For a better understanding of the situation, we strip the Dirac equation to its bare bones. After diagonalizing it, we find that the Lagrangian consists of elementary units of the form

$$L = \overline{\psi}(i\partial_t \psi - M\psi) ; \qquad p_\psi = i\overline{\psi} ; \qquad H = \overline{\psi} M \psi . \tag{3.12}$$

If we were using ordinary numbers, the only way to obtain a lower bound on H would be by identifying $\overline{\psi}$ with ψ. Then, however, the kinetic part of the Lagrangian would become a time-derivative:

$$\overline{\psi}\partial_t\psi \to \tfrac{1}{2}\partial_t(\overline{\psi}\,\psi) , \tag{3.13}$$

so that it could not contribute to the action. One concludes that, only in the space of *anticommuting numbers*, can the Lagrangian (3.12) make sense. Thus, one replaces the Poisson brackets for ψ and $\overline{\psi}$ by *anti*commutators:

$$\{\overline{\psi},\,\psi\} \equiv \overline{\psi}\,\psi + \psi\,\overline{\psi} = 1\;;\qquad \{\psi,\,\psi\} = 0\;;\qquad \{\overline{\psi},\,\overline{\psi}\} = 0\;. \tag{3.14}$$

The elementary representation of this algebra is in a 'Hilbert space' consisting of just two states (the empty state and the one-particle state), in which the operators ψ and $\overline{\psi}$ act as annihilators and creators:

$$\psi = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\;;\qquad \overline{\psi} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\;;\qquad H = \begin{pmatrix} 0 & 0 \\ 0 & M \end{pmatrix}\;. \tag{3.15}$$

Returning to the non-diagonalized case, we can keep the Lagrangian (3.9) and Hamiltonian (3.11) when the commutation rules (3.14) are replaced by

$$\begin{aligned} \{\overline{\psi}^{i}(\mathbf{x}),\,\psi_{j}(\mathbf{x}')\} &= \delta^{i}_{j}\,\delta(\mathbf{x}-\mathbf{x}')\;; \\ \{\psi_{i}(\mathbf{x}),\,\psi_{j}(\mathbf{x}')\} &= 0\;;\qquad \{\overline{\psi}^{i}(\mathbf{x}),\,\overline{\psi}^{j}(\mathbf{x}')\} = 0\;. \end{aligned} \tag{3.16}$$

The commutation rules (3.16) turn Dirac particles into fermions. It appears to be a condition for any Lorentz-invariant quantum theory to be consistent, that *integer* spin particles must be *bosons* and particles whose spin is an integer $+\frac{1}{2}$ must be *fermions*.

3.3.   The path integral for anticommuting fields

Let us now extend the notion of path integrals to include Dirac fields. This means we have to integrate over anticommuting numbers, to be called θ_i, where i is some index (possibly including $\mathbf{x}$). They are numbers, not operators, so all anticommutators vanish. Consider the Taylor expansion of a function of a variable θ. Since $\theta^2 = 0$, this expansion has only two coefficients:

$$f(\theta) = f(0) + f'(0)\theta\;. \tag{3.17}$$

So, this is the most general function of θ that one can have. It is generally agreed that one should define integrals for anticommuting numbers θ by postulating

$$\int \mathrm{d}\theta\,1 \equiv 0\;;\qquad \mathrm{d}\theta\,\theta \equiv 1\;. \tag{3.18}$$

The reason for this definition is that one can manipulate these expressions in the same way as integrals over ordinary numbers:

$$\int \mathrm{d}\theta\,f(\theta+\alpha) = \int \mathrm{d}\theta\,f(\theta)\;;\qquad \int \mathrm{d}\theta\frac{\partial f(\theta)}{\partial \theta} = 0\;, \tag{3.19}$$

22

etc.

Now, consider the Hamiltonian for just one fermionic degree of freedom, (3.15), which we write as

$$H = M\, b^\dagger b\; ; \qquad \{b,\, b^\dagger\} = 1\; ; \qquad b^2 = (b^\dagger)^2 = 0\; , \tag{3.20}$$

and a wave function $\psi = \begin{pmatrix} \psi_0 \\ \psi_1 \end{pmatrix}$. Define the following function of θ:

$$\psi(\theta) \equiv \psi_0 \theta + \psi_1\; , \tag{3.21}$$

This now serves as our wave function. It is not hard to derive how the annihilation operator b and the creation operator $b^\dagger$ act on these wave functions:

$$\text{if} \qquad \phi = b\,\psi \qquad \text{then} \qquad \phi(\theta) = \theta\,\psi(\theta)\; , \tag{3.22}$$

or:

$$b = \theta\; ; \qquad b^\dagger = \frac{\partial}{\partial \theta}\; . \tag{3.23}$$

We now wish to express the evolution of a fermionic wave function in terms of a path integral, just as in subsection 2.4. Consider a short time interval δt. Then, ignoring all terms of order $(\delta t)^2$, one derives

$$
\begin{aligned}
e^{-i\delta t\, H}\psi(\theta_1) &= \psi_0\,\theta_1 + (1 - iM\,\delta t)\psi_1 \\
&= \int d\theta_0(-\theta_1 + \theta_0 - iM\delta t\theta_0)(\psi_0\theta_0 + \psi_1) \\
&= \int d\theta_0 \int d\bar\theta \Big(1 + \bar\theta(-\theta_1 + \theta_0 - iM\delta t\theta_0)\Big)\psi(\theta_0) \\
&= \int d\theta_0 \int d\bar\theta\, e^{\bar\theta(-\theta_1 + \theta_0 - iM\delta t\theta_0)}\psi(\theta_0)\; .
\end{aligned}
\tag{3.24}
$$

Repeating this procedure over many infinitesimal time intervals, with $T = N\,\delta t$, one arrives at the formal expression

$$\psi(\theta_T) = \int d\bar\theta_{T-1}d\theta_{T-1} \;\cdots\; d\bar\theta_0 d\theta_0 \;\exp \sum_{\tau=0}^{N-1} \delta t\Big(\bar\theta_\tau(\frac{-\theta_{\tau+1} + \theta_\tau}{\delta t} - iM\theta_\tau)\Big)\,\psi(\theta_0)\; . \tag{3.25}$$

The exponential tends to

$$i \int dt\, L(t)\; . \tag{3.26}$$

Thus, as in the bosonic case, the evolution operator is formally the path integral of e^{iS} over all (anticommuting) fields $\psi_i(\mathbf{x}, t)$, where the action S is the time integral[4]. of the Lagrangian L, and indeed the space-time integral of the Lagrange density $\mathcal{L}(\mathbf{x}, t)$.

---

[4]In some applications, careful considerations of the boundary conditions for Dirac's equation, require an extra boundary term to be added to the action (3.25)

3.4. The Feynman rules for Dirac fields

Let M_{ij} be any matrix that can be diagonalized. Using Eqs. (3.18), we find the integral

$$\prod_i \int \mathrm{d}\theta_i \int \mathrm{d}\bar{\theta}_i \, e^{\bar{\theta}_i M_{ij} \theta_j} = \det_{ij}(M) \ , \tag{3.27}$$

which can be easily checked by diagonalizing M, and writing

$$\int \mathrm{d}\theta \int \mathrm{d}\bar{\theta} \, e^{\bar{\theta} M \theta} = \int \mathrm{d}\theta \int \mathrm{d}\bar{\theta} \, (1 + \bar{\theta} M \theta) = M \ . \tag{3.28}$$

Thus, a Gaussian integral over anticommuting numbers gives a result very similar to that over commuting numbers, except that we get $\det(M)$ rather than $C/\det(M)$. Writing

$$
\begin{aligned}
M &= M_0 + \delta M \ ; \\
\det(M) &= e^{\mathrm{Tr}(\log M)} = 1 + \mathrm{Tr}\,(\log M) + \tfrac{1}{2}(\mathrm{Tr}\,\log M)^2 + \cdots \ ,
\end{aligned}
\tag{3.29}
$$

we see that this can be obtained from $\det(M^{-1})$ by switching the signs of all odd terms in this expansion. Since the Nth term corresponds to a Feynman diagram with N closed fermionic loops, one derives that the Feynman rules can be read off from the ones for ordinary commuting fields, by switching a sign whenever a closed fermionic loop is encountered.

We have

$$-\mathrm{Tr}\,\log M = -\mathrm{Tr}\,\log M_0 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \mathrm{Tr}\,(M_0^{-1} \delta M)^n \ . \tag{3.30}$$

Here, as in the bosonic case, $-M_0$ is the propagator of the theory, and δM represents the contribution from any perturbation. Thus, if our Lagrangian, including possible interaction terms, is

$$\mathcal{L} = -\overline{\psi}_i (m_{(i)} + \gamma^\mu \partial_\mu) \psi_i + \overline{\psi}_i g_{ij}(\phi) \psi_j \ , \tag{3.31}$$

then the propagator, in Fourier space, is

$$(m_{(i)} + i\gamma^\mu k_\mu)^{-1} = \frac{m_{(i)} - i\gamma^\mu k_\mu}{m_{(i)}^2 + k^2 - i\varepsilon} \ , \tag{3.32}$$

while $g_{ij}(\phi)$ generates the interaction vertices of a Feynman diagram. The $i\varepsilon$ term is chosen as in bosonic theories, for the same reason as there: the vacuum state must be the state with lowest energy.

The poles in the propagator can be used to define in- and out-going particles, by adding source terms to the Lagrangian:

$$\delta \mathcal{L} = \overline{\eta}(x)\psi + \overline{\psi}\eta(x) \ , \tag{3.33}$$

24

where $\eta(x)$ and $\overline{\eta}(x)$ are kept fixed, as anticommuting numbers. We could proceed to derive the precise rules for in- and out-going particles with spin up or down, but it is more convenient to postpone this until we discuss the unitarity property of the S-matrix, where these rules are required explicitly, and where we find the precise prescription for the normalization of these states (section 6).

Note that our Lagrangian is always kept to be bilinear in the anticommuting fields. This is because we insist that $\mathcal{L}$ itself must be a commuting number and, furthermore, terms that are *quartic* in the fermionic fields have too high a dimension. We will see in section 7 why such terms have to be avoided.

4. Gauge fields

We continue to search for *elementary* fields, whose Lorentz covariant field equations can be subject to our quantization program. In principle, such fields could come as any arbitrary representation of the Poincaré algebra, that is, we might consider any kind of tensor field, $A_{\mu\nu\lambda\cdots}(\mathbf{x}, t)$. It turns out, however, that tensors with more than one Lorentz index cannot be used. This is because we wish the *energy density* of a field to be bounded from below, and in addition, we wish the *dimensionality* of the interactions to be sufficiently low, such that all coupling strengths have mass dimension zero or positive. Coupling strengths with mass dimension less than zero give rise to unacceptably divergent expressions for the contributions of the interactions at short scales. A prime example of a field one would *like* to include is the gravitational field described by the metric $g_{\mu\nu}(x)$, but its only possible interaction is the gravitational one, whose coupling strength, Newton's constant G_N, has the wrong dimension. The non-renormalizable theories one then obtains are the subject of intense investigations but fall outside the scope of this paper (see C. Rovelli's contribution in this book).

So, only spin-one fields $A_\mu^a(x)$ are left for consideration. Here, μ is a Lorentz index, while the number of field types is counted by the index $a = 1, \cdots, N_V$. These fields should describe the creation and annihilation of spin-one particles. When at rest, such a particle will be in one of three possible spin states. Yet, to be Lorentz-invariant, a vector field A_μ should have four components. One of these, at least, should be unphysical, although one might think of accepting an extra, spinless particle to be associated to the vector particles. More important therefore is the consideration that, in the corresponding classical theory, the energy should be bounded from below.

This then rules out the treatment of a four-vector field as if we had four scalar fields, because the Lorentz-invariant product has an indefinite metric. Can we construct a Lagrangian for a vector field that gives a Hamiltonian that is bounded from below?

Let us look at the high-momentum limit for one of these vector fields. The only two

terms in a Lagrangian that can survive there are:

$$\mathcal{L} = -\tfrac{1}{2}\alpha\,(\partial_\mu A_\nu)^2 + \tfrac{1}{2}\beta\,\partial_\mu A_\mu\,\partial_\nu A_\nu \ , \tag{4.1}$$

since other terms of this dimensionality can be reduced to these ones by partial integration of the action, while mass terms (terms without partial derivatives) become insignificant. We have for the canonical momentum fields

$$
\begin{aligned}
E^i &= \frac{\partial \mathcal{L}}{\partial \partial_0 A_i} &= \alpha\,\partial_0 A_i \qquad (i=1,2,3); \\
E^0 &= \frac{\partial \mathcal{L}}{\partial \partial_0 A_0} &= (\beta-\alpha)\,\partial_0 A_0 - \beta\,\partial_i A_i \ .
\end{aligned}
\tag{4.2}
$$

Now, consider the Hamiltonian density $\mathcal{H} = E^\mu\,\partial_0 A_\mu - \mathcal{L}$. It must be bounded from below for all field configurations $A_\mu(\mathbf{x},t)$. Let us first consider the case when the spacelike components A_i and all spacelike derivatives ∂_i are negligible compared to $\partial_0 A_0$:

$$\mathcal{H} \to \tfrac{1}{2}(\beta-\alpha)(\partial_0 A_0)^2 \ , \tag{4.3}$$

then, when A_0 and all time-derivatives are negligible:

$$\mathcal{H} \to \tfrac{1}{2}\alpha(\partial_i A_j)^2 - \tfrac{1}{2}\beta(\partial_i A_i)^2 \ . \tag{4.4}$$

These must all be bounded from below. Eq. (4.3) dictates that $\beta \geq \alpha$, while Eq. (4.4) dictates that $\alpha \geq \beta$. We conclude that $\alpha = \beta$, which we can both normalize to one. Since total derivatives in the Lagrangian do not count, we can then rewrite the original Lagrangian (4.1) as

$$\mathcal{L} \to -\tfrac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} \ , \qquad F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu \ . \tag{4.5}$$

Realizing that this is the Lagrangian for ordinary QED, we know that its energy-density is properly bounded from below. We conclude that *every* vector field theory must have a Lagrangian that approaches Eq. (4.5) at high energies and momenta.

We do note, that with the choice $\alpha = \beta$, both (4.3) and (4.4) tend to zero. Indeed, any field A^a_μ that can be written as a space-time gradient, $A^a_\mu = \partial_\mu\Lambda^a(\mathbf{x},t)$, has $F^a_{\mu\nu} = 0$, and hence contributes neither to the Lagrangian nor to the Hamiltonian. Such fields could be arbitrarily strong, yet carry zero energy. They would represent particles and forces without energy. This is unacceptable in a decent Quantum Field Theory. How do we protect our theory against such features?

There is exactly one way to do this. We must make sure that field replacements of the type

$$A^a_\mu \to A^a_\mu + \partial_\mu\Lambda^a(x) + \cdots \ , \tag{4.6}$$

26

do not affect at all the physical state that we are describing. This is what we call a local gauge transformation. We must insist that our theory is invariant under local gauge transformations. The ellipses in Eq. (4.6) indicate that we allow extra terms that do not contribute to the bilinear part of the Lagrangian (4.5). Thus, we arrive at Yang-Mills field theory.

4.1. The Yang-Mills equations

[3] Our conclusion from the above is that every vector field is associated to a *local gauge symmetry*. The dimensionality of the local gauge group must be equal to N_V, the number of vector fields present. Besides the vector fields, the local symmetry transformations may also affect the scalar and spinor fields. In short, the vector fields must be Yang-Mills fields. We here give a brief summary of Yang-Mills theory.

We have a local Lie group with elements $\Omega(x)$ at the point x. Let the matrices T^a, $a = 1, \cdots, N_V$ be its infinitesimal generators:

$$\Omega(x) = \mathbb{I} + i \sum_a \Lambda^a(x) T^a ; \qquad T^a = (T^a)^\dagger . \tag{4.7}$$

Characteristic for the group are its structure constants f_{abc}:

$$[T^a, T^b] = i f_{abc} T^c . \tag{4.8}$$

As is well-known in group theory, one can choose the normalization of T^a in such a way that the f_{abc} are totally antisymmetric:

$$f_{abc} = -f_{bac} = f_{bca} . \tag{4.9}$$

Usually, the spinor fields $\psi(x)$ and scalar fields $\phi(x)$ are introduced in such a way that they transform as (sets of irreducible) *representations* of the gauge group. A *local gauge transformation* is then:

$$\psi'(x) = \Omega(x)\psi(x) ; \qquad \phi'(x) = \Omega(x)\phi(x) , \tag{4.10}$$

and in infinitesimal form:

$$\psi'(x) = \psi(x) + i\Lambda^a(x)T^a\psi(x) + \mathcal{O}(\Lambda)^2 , \tag{4.11}$$

and similarly for $\phi(x)$. The *dimension* of the irreducible representation can be different for different field types. So, scalar and spinor fields usually form gauge-vectors of various dimensionalities.

Our vector fields $A_\mu^a(x)$ are most conveniently introduced by demanding the possibility of constructing gauge-covariant gradients of these fields:

$$D_\mu\psi(x) \equiv (\partial_\mu + igA_\mu^a(x)T^a)\psi(x) , \tag{4.12}$$

where g is a freely adjustable coupling parameter. The repeated indices a, denoting the different species of vector fields, are to be summed over. By demanding the transformation rule

$$(D_\mu \psi(x))' \; = \; \Omega(x) D_\mu \psi(x) \; = \; D_\mu \psi(x) + i\Lambda^a(x) T^a D_\mu \psi(x) + \mathcal{O}(\Lambda)^2 \; , \qquad (4.13)$$

one easily derives the transformation rule for the vector fields $A_\mu^a(x)$:

$$\begin{aligned} igA_\mu^{a\prime}(x) T^a \; &= \; \Omega(x) \Big(\partial_\mu + igA_\mu^a(x) T^a \Big) \Omega^{-1}(x) \\ &= \; igA_\mu^a(x) T^a - i\partial_\mu \Lambda^a(x) T^a + g[T^a, T^b] \Lambda^a(x) A_\mu^b(x) \qquad (4.14) \end{aligned}$$

(omitting the $\mathcal{O}(\Lambda)^2$ terms). With Eq. (4.8), this becomes

$$A_\mu^{a\prime}(x) = A_\mu^a(x) - \tfrac{1}{g} \partial_\mu \Lambda^a(x) + f_{abc} \Lambda^b(x) A_\mu^c(x) \; . \qquad (4.15)$$

If we ensure that all gradients used are covariant gradients, we can directly construct the general expressions for Lagrangians for scalar and spinor fields that are locally gauge-invariant:

$$\begin{aligned} \mathcal{L}_{\text{scalar}}^{\text{inv}}(x) \; &= \; -\tfrac{1}{2}(D_\mu \phi)^2 - V(\phi^2) \; ; & (4.16) \\ \mathcal{L}_{\text{Dirac}}^{\text{inv}}(x) \; &= \; -\overline{\psi}(\gamma^\mu D_\mu + m)\psi \; , & (4.17) \end{aligned}$$

and in addition other possible invariant local interaction terms without derivatives.

The commutator of two covariant derivatives is

$$\begin{aligned} [D_\mu, D_\nu]\psi(x) \; &= \; igF_{\mu\nu}^a(x) T^a \psi(x) \; ; \\ F_{\mu\nu}^a(x) \; &= \; \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf_{abc} A_\mu^b A_\nu^c \; . \qquad (4.18) \end{aligned}$$

Unlike $A_\mu^a(x)$ or the direct gradients of $A_\mu^a(x)$, this *Yang-Mills field* $F_{\mu\nu}^a$ transforms as a true adjoint representation of the local gauge group:

$$F_{\mu\nu}^a{}'(x) = F_{\mu\nu}^a(x) + f_{abc} \Lambda^b(x) F_{\mu\nu}^c(x) \; . \qquad (4.19)$$

This allows us to construct a locally gauge invariant Lagrangian for the vector field:

$$\mathcal{L}_{\text{YM}}^{\text{inv}}(x) = -\tfrac{1}{4} F_{\mu\nu}^a(x) F_{\mu\nu}^a(x) \; . \qquad (4.20)$$

The structure constants f_{abc} in the definition 4.18 of the field $F_{\mu\nu}$ implies the presence of interaction terms in the Yang-Mills Lagrangian 4.20. If f_{abc} is non-vanishing, we talk of a *non-Abelian* gauge theory.

28

There is one important complication in the case of fermions: the Dirac matrix $\gamma^5 \equiv \gamma^1 \gamma^2 \gamma^3 \gamma^4$ can be used to project out the *chiral sectors*:

$$\psi \equiv \psi_L + \psi_R \; ; \qquad \psi_L = \tfrac{1}{2}(1 + \gamma^5)\psi \; ; \quad \psi_R = \tfrac{1}{2}(1 - \gamma^5)\psi \; . \qquad (4.21)$$

Since the kinetic part of a Dirac Lagrangian can be split according to

$$\mathcal{L}_{\text{Dirac}} = -\overline{\psi}_L(\gamma D)\psi_L - \overline{\psi}_R(\gamma D)\psi_R \; , \qquad (4.22)$$

we may choose the left-handed fields ψ_L to be in representations different from the right-handed ones, ψ_R. However, since a mass term joins left to right:

$$-m\overline{\psi}\,\psi = -m\overline{\psi}_L\psi_R - m\overline{\psi}_R\psi_L \; , \qquad (4.23)$$

such terms would then be forbidden, hence such chiral fields must be massless. Secondly, not all combinations of chiral fermions are allowed. An important restriction is discussed in section 8. The fields ψ_L turn out to describe spin-$\tfrac{1}{2}$ massless particles with only the left-rotating helicity, while their antiparticles, described by $\overline{\psi}_L$, have only the right-rotating helicity.

4.2. The need for local gauge-invariance

In the early days of Gauge Theory, it was thought that local gauge-invariance could be an 'approximate' symmetry. Perhaps one could add mass terms for the vector field that violate local symmetry, but make the model look more like the observed situation in particle physics. We now know, however, that such models suffer from a serious defect: they are non-renormalizable. The reason is that renormalizability requires our theory to be consistent up to the very tiniest distance scales. A mass term would, at least in principle, turn the field configurations described by the $\Lambda(x)$ contributions in Eq. (4.15) into physically observable fields (the Lagrangian now does depend on $\Lambda(x)$). But, since the kinetic term for $\Lambda(x)$ is lacking, violently oscillating Λ fields carry no sizeable amount of energy, so they would not be properly suppressed by energy conservation. Uncontrolled short distance oscillations are the real, physical cause for a theory being non-renormalizable.

It is similar uncontrolled short-distance fluctuations of the space-time metric that cause the quantized version of General Relativity ("Quantum Gravity") to be non-renormalizable. Drastic measures (String Theory?) are needed to repair such a theory.

Since renormalizability provides the required coherence of our theories, local gauge symmetry, described by Eqs. (4.10) and (4.15), *must* be an exact, not an approximate symmetry of any Quantum Field Theory.[5] Obviously, the fact that most vector particles

---

[5]One apparent exception could be the case where the longitudinal component decouples completely, which happens in massive QED. But even in that case, it is better to view the longitudinal photon as a Higgs field, see section 5.

in the sub-atomic world do carry mass must be explained in some other way. It is here that the Brout-Englert-Higgs mechanism comes to the rescue, see Section 5.

4.3. Gauge fixing

The longitudinal parts of the vector fields do not occur directly in the Yang-Mills Lagrangian (4.20), exactly because of its invariance under transformations of the form (4.15). Yet if we wish to describe solutions, we need to choose a longitudinal component. This is why we wish to impose some additional constraint, the so-called gauge condition, on our description of the solutions, both in the classical and in the quantized theory. In electrodynamics, we usually impose a constraint such as $\partial_\mu A_\mu(x) = 0$ or $A_0 = 0$. In a Yang-Mills theory, such a constraint is needed for each value of the index a. A gauge fixing term is indicated by a field $C^a(x)$ which is put equal to zero:

$$C^a(x) = 0 ; \qquad a = 1, \cdots, N_V ; \qquad \text{where} \qquad (4.24)$$
$$\text{either} \qquad C^a(x) = \partial_\mu A^a_\mu(x) \qquad \text{(Feynman gauge)}, \qquad (4.25)$$
$$\text{or} \qquad C^a(x) = A^a_0(x) \qquad \text{(timelike gauge)}, \qquad (4.26)$$

or other possible gauge choices. It is always possible to find a $\Lambda^a(x)$ that obeys such a condition. For instance, to obtain the Feynman gauge (4.25), all one has to do is extremize an integral under variations of the gauge group:

$$\delta \int \mathrm{d}^4x \left(A^a_\mu(x) \right)^2 = 0 \quad \rightarrow \quad \partial_\mu A^a_\mu(x) = 0 . \qquad (4.27)$$

For the classical theory, the most elegant way to impose such a gauge condition is by adding a Lagrange multiplier term to the Lagrangian:

$$\mathcal{L}(x) = \mathcal{L}^{\text{inv}}(x) + \lambda^a(x)C^a(x) , \qquad (4.28)$$

where $C^a(x)$ is any of the possible gauge fixing terms and $\lambda^a(x)$ a free kinematical variable. Here, $\mathcal{L}^{\text{inv}}$ stands for the collection of all gauge-invariant terms in the Lagrangian. The Euler-Lagrange equations of the theory then automatically yield the Yang-Mills field equations plus the constraint, apart from a minor detail: the boundary condition. Varying the gauge transformations, one finds, since $\mathcal{L}^{\text{inv}}$ does not vary, $D_\mu \lambda^a(x) = 0$. We need to impose the stricter equation $\lambda^a(x) = 0$, which is obtained by imposing $\lambda^a(x) = 0$ at the boundaries of our system.

Alternatively, one can replace the invariant Lagrangian by

$$\mathcal{L}(x) = \mathcal{L}^{\text{inv}}(x) - \tfrac{1}{2}\left(C^a(x) \right)^2 , \qquad (4.29)$$

which has the advantage that, after partial integration, the bilinear part becomes very simple: $\mathcal{L} = \tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - \tfrac{1}{2}(\partial_\mu A^\mu)^2 \rightarrow -\tfrac{1}{2}(\partial_\mu A_\nu)^2$, so that the vector field can be treated

30

as if it were just 4 scalars. Again, varying the gauge transformation $\Lambda^a(x)$, one finds $D_\mu C^a(x) = 0$, which must be replaced by the more stringent condition $C^a(x) = 0$ by adding the appropriate boundary condition.

Note that the Lagrange-Hamilton formalism could give the wrong sign to the energy of some field components; we should continue to use the energy deduced before imposing the gauge constraint. If we use the timelike gauge (4.26), the energy is correct, but the theory appears to lack Lorentz invariance. Lorentz transformations must now be accompanied by gauge transformations.

How is the gauge constraint to be handled in the quantized theory? This problem was solved by B.S. DeWitt[4],[5] and by Faddeev and Popov[6]. The gauge constraint is to be imposed in the integrand of the functional integral:

$$Z = \int \mathcal{D}A(x) \int \mathcal{D}\phi(x) \cdots e^{i \int \mathrm{d}^4x \mathcal{L}^{\mathrm{inv}}(x)} \prod_{a,x} \delta(C^a(x)) \Delta\{A, \phi\} \; . \tag{4.30}$$

Thus, we integrate only over those field configurations that obey the gauge condition. $\Delta\{A, \phi\}$ is a Jacobian factor, which we will discuss in a moment. The formal delta function can be replaced by a Lagrange multiplier:

$$\int \mathcal{D}\lambda^a(x) e^{i \int \mathrm{d}^4x \, \lambda^a(x) C^a(x)} \; , \tag{4.31}$$

and indeed, if $\lambda^a(x)$ is simply added to the list of dynamical field variables of the theory, the Feynman rules can be derived unambiguously as they were for the scalar and the spinor case.

There is, however, a problem. It appears to be difficult to prove gauge-invariance. More precisely: we need to ascertain that, if we make the transition to a different gauge fixing function $C^a(x)$, the physical contents of the theory, in particular the scattering matrix, remains the same. The difficulty has to do with the *measure* of the integral. It is not gauge-invariant, unless we add the extra term $\Delta\{A, \phi\}$ in Eq. (4.30). This term is associated to the volume of an infinitesimal gauge transformation. Suppose that the field combination $C^a(x)$ transforms under a gauge transformation as

$$C^{a\prime}(x) = C^a(x) + \frac{\partial C^a(x)}{\partial \Lambda^b(x')} \Lambda^b(x') \; , \tag{4.32}$$

then the required volume term is the Jacobian

$$\Delta\{A, \phi\} = \det \left(\frac{\partial C^a(x)}{\partial \Lambda^b(x')} \right) \; . \tag{4.33}$$

The determinant is computed elegantly by using the observation in subsection 3.4 that that integral over anticommuting variables gives a determinant (Eq. (3.27)). So, we

introduce anticommuting scalar fields η and $\overline{\eta}$, and then write

$$(4.33) = \int \mathcal{D}\eta^a(x) \int \mathcal{D}\overline{\eta}^a(x) \, \exp\left(\overline{\eta}^a(x)\frac{\partial C^a(x)}{\partial \Lambda^b(x')}\eta^b(x')\right) . \tag{4.34}$$

This is called the *Faddeev-Popov* term in the action. Taking everything together, we arrive at the following action for a Yang-Mills theory:

$$\mathcal{L}(x) = \mathcal{L}^{\mathrm{inv}}(x) + \lambda^a(x)C^a(x) + \overline{\eta}^a(x)\frac{\partial C^a(x)}{\partial \Lambda^b(x')}\eta^b(x') . \tag{4.35}$$

It is also possible to find the quantum analogue for the classical Lagrangian (4.29). First, replace $C^a(x)$ by $C^a(x) - F^a(x)$, where $F^a(x)$ is a fixed but x-dependent quantity in the functional integral (4.35). Physical effects should be completely independent of $F^a(x)$. Therefore, we can functionally integrate over $F^a(x)$, using any weight factor we like. Choose the weight factor $e^{-\frac{1}{2}\int \mathrm{d}^4 x(F^a(x))^2}$. The Lagrange multiplier $\lambda^a(x)$ now simply forces $C^a(x)$ to be equal to $F^a(x)$. We end up with the effective Lagrangian[6]

$$\mathcal{L}(x) = \mathcal{L}^{\mathrm{inv}}(x) - \tfrac{1}{2}\Big(C^a(x)\Big)^2 + \overline{\eta}^a(x)\frac{\partial C^a(x)}{\partial \Lambda^b(x')}\eta^b(x') . \tag{4.36}$$

This is the most frequently used Lagrangian for gauge theories. In contrast to the Lagrangians for scalar and spinor fields, not all fields here represent physical particles. The longitudunal part of the vector fields, and the fermionic yet scalar fields $\overline{\eta}$ and η are "ghosts".

4.4. Feynman rules

The Feynman rules, needed for the computation of the scattering matrix elements using perturbation theory, can be read off directly from the gauge-fixed Lagrangian (4.35) or (4.36). In both cases, we first split off the bilinear parts[7], writing the Lagrangian as

$$\mathcal{L} = -A_\alpha(x)\hat{M}_{\alpha\beta}A_\beta(x) - \overline{\psi}_\alpha(x)\hat{D}_{\alpha\beta}\psi_\beta(x) + \mathcal{L}^{\mathrm{int}} , \tag{4.37}$$

where $\mathcal{L}^{\mathrm{int}}$ contains all trilinear and quadrilinear terms. Here, $A_\alpha(x)$ is short for all bosonic (scalar and vector) fields, and $\overline{\psi}$ and ψ for both the Dirac fermions and the Faddeev-Popov fermions. The coefficients $\hat{M}_{\alpha\beta}$, $\hat{D}_{\alpha\beta}$ and the trilinear coefficients may contain the gradient operator $\partial/\partial x^\mu$. After Fourier expansion, this will turn into a factor ik_μ.

---

[6]One usually absorbs the factor $1/g$ of Eq. (4.15) into the definition of the $\overline{\eta}$ field.

[7]One may decide to leave small corrections to the bilinear parts of the Lagrangian to be treated together with the higher order terms as if they were 'two-point vertices'.

— The *propagators* $\hat{P}_{\alpha\beta}$ and $\hat{P}^{\text{ferm}}_{\alpha\beta}$ will be the inverse of the coefficients $\hat{M} - i\varepsilon$ and $\hat{D} - i\varepsilon$, so, for instance

$$\text{if} \quad \hat{M}_{\alpha\beta} = (m_{(\alpha)} - \partial^2_\mu)\delta_{\alpha\beta} \quad \text{then} \quad \hat{P}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{m^2_{(\alpha)} + k^2 - i\varepsilon} \; ;$$

$$\text{if} \quad \hat{D}_{\alpha\beta} = (m_{(\alpha)} + \gamma^\mu\partial_\mu)\delta_{\alpha\beta} \quad \text{then} \quad \hat{P}^{\text{ferm}}_{\alpha\beta} = \frac{(m_{(\alpha)} - i\gamma^\mu k_\mu)\delta_{\alpha\beta}}{m^2_{(\alpha)} + k^2 - i\varepsilon} \; . \quad (4.38)$$

— The vertices are generated by the trilinear and quadrilinear terms of $\mathcal{L}^{\text{int}}$, just as in subsection 2.5. If we have source terms such as $J^a(x)\phi_a(x)$, $\overline{\eta}^i(x)\psi_i(x)$ or $\overline{\psi}^i(x)\eta_i(x)$, then these correspond to propagators ending into points, where the momentum k has to match a given Fourier component of the source. All this can be read off neatly from formal expansions of the functional integral such as (2.46).

— There is an overall minus sign for every fermionic closed loop.

— Every diagram comes with canonical coefficients such as $1/k!$ and $(2\pi)^{-4N}$ where $k!$ is the dimension of the diagram's internal symmetry group, and N counts the number of loop integrations. These coefficients can be obtained by comparing functional integrals with ordinary integrals.

— There is a normalization coefficient for every external line, depending on the wave function chosen for the in- and out-going particles. We return to this in section 6.

Note that any terms in the Lagrangian that can be written as a gradient of some (locally defined) field configuration can be replaced by zero. This is because (under sufficiently carefully chosen boundary conditions) such terms do not contribute to the total action $S = \int \mathrm{d}^4x \mathcal{L}(x)$.

4.5. BRST symmetry

As the reader may have noted, we departed from our original intention, to keep space and time on a lattice and only turn to the continuum limit at the very end of a calculation. We have not even started doing calculations, and already the Feynman rules were formulated as if the fields lived on a space-time continuum. Indeed, we should have kept space and time discrete, so that the functional integral is nothing but an ordinary integral in a space with very many, but still a finite number of, dimensions. In practice, however, the continuum is a lot easier to handle, so, often we do not explicitly mention the finite size meshes of space and time.

Our first attempt to formulate the continuum limit will be in section 7. We will then see that the coefficients in the Lagrangian (4.36) have to be *renormalized*. The following question then comes up:

If we see a Lagrangian that looks like (4.36), how can we check that its coefficients are those of a genuine gauge theory?

The answer to this question is that the gauge-fixed Lagrangians (4.35) and (4.36) possess a *symmetry*. The first attempts to identify the symmetry in question gave negative results, because the ghost field is fermionic while the gauge fixing terms are bosonic. In the early days we thought that the required relation between the gauge fixing terms and the ghost terms had to be checked by inspection.[15] But the complete answer was discovered by Becchi, Rouet and Stora[7], and independently by Tyutin[8]. The symmetry, called BRST symmetry, is a *supersymmetry*. For the Lagrangian (4.36), which is slightly more general than (4.35), the transformation rules are

$$
\begin{array}{rcll}
A'_\alpha(x) & = & A_\alpha(x) + \overline{\varepsilon}\, \frac{\partial A_\alpha(x)}{\partial \Lambda^b(x')}\, \eta^b(x') \; ; & (a) \\[2mm]
\eta^{a\prime}(x) & = & \eta^a(x) + \tfrac{1}{2}\overline{\varepsilon}\, f_{abc}\eta^b(x)\eta^c(x) \; ; & (b) \\[2mm]
\overline{\eta}^{a\prime}(x) & = & \overline{\eta}^a(x) + \overline{\varepsilon}\, C^a(x) \; , & (c)
\end{array}
\qquad (4.39)
$$

where the anticommuting number $\overline{\varepsilon}$ is the infinitesimal generator of this (global) supersymmetry transformation.

The invariance of the Lagrangian (4.36) under this supersymmetry transformation is easy to check, except perhaps the cancellation of the variation of the last term against the contribution of (4.39 b):

$$
\overline{\eta}^a\, \frac{\partial C^a}{\partial \Lambda^b}\, \tfrac{1}{2}\overline{\varepsilon}\, f_{bcd}\eta^c\eta^d + \overline{\eta}^a\, \frac{\partial}{\partial \Lambda^c}\frac{\partial C^a}{\partial \Lambda^d}\, \eta^c\eta^d \;\; = \cdots \; .
\qquad (4.40)
$$

Substituting some practical examples for the gauge constraint function C^a, one discovers that these terms always cancel out. The reason for (4.40) to vanish is the fact that gauge transformations form a group, implying the Jacobi identity:

$$
f_{asb}f_{scd} + f_{asc}f_{sdb} + f_{asd}f_{sbc} = 0 \; .
\qquad (4.41)
$$

The converse is more difficult to prove: *If* a theory is invariant under a transformation of the form (4.39) (BRST invariance), then it is a gauge-fixed local gauge theory. What is really needed in practice, is to show that the ghost particles do not contribute to the S-matrix. This indeed follows from BRST invariance, via the so-called Slavnov-Taylor identities[16][17], relations between amplitudes that follow from this symmetry.

5. The Brout-Englert-Higgs mechanism

The way it is described above, Yang-Mills gauge theory does not appear to be suitable to describe *massive* particles with spin one. However, in our approach we concentrated only on the high-energy, high-momentum limit of theories for vector particles, by assuming the Lagrangian to take the form (4.1) there. Mass terms dominate in the infra-red, or low energy domain. Here, one may note that we have not yet exploited all possibilities.

We need to impose exact local gauge-invariance, as explained in subsection 4.2. So our theory must be constructed along the lines expounded in subsection 4.1. All scalar and spinor fields must come as representations of the gauge group. So, what did we overlook?

In our description of the most general, locally gauge-invariant Lagrangian, it was tacitly assumed that the minimum of the scalar potential function $V(\phi)$ occurs at $\phi = 0$, so that, as one may have in global symmetries, the symmetry is evident in the particle spectrum: physical particles come as representations of the full local symmetry group. But, as we have seen in the case of a global symmetry, in subsection 2.2, the minimum of the potential may occur at other values of ϕ. If these values are not invariant under the gauge group, then they form a non-trivial representation of the group, invariant only under a subgroup of the gauge group. It is the invariant subgroup, if at all non-trivial, of which the physical particles will form representations, but the rest of the symmetry is hidden. Indeed, if we switch off the coupling to the vector fields, we obtain again the situation described in subsection 2.2. As was emphasized there, the particle spectrum then contains massless particles, the Goldstone bosons. These Goldstone bosons represent the field excitations associated to a global symmetry transformation, which does not affect the energy: hence the absence of mass.

But, Global gauge Goldstone bosons *do* carry a kinetic term. Therefore, they do carry away energy when moving with the speed of light. This is because a global symmetry only dictates the Goldstone field to carry no energy if the field is space-time independent.

In contrast, *local* gauge symmetries demand that Goldstone fields also carry no energy when they do depend on space and time. In the case of a local symmetry, therefore, Goldstone modes are entirely in the ghost sector of the theory; Goldstone particles then are unphysical. Let us see how this happens in an example.

5.1. The SO(3) case

As a prototype, we take the group $SO(3)$ as our local gauge group, and for simplicity we ignore the contributions of loop diagrams, which represent the higher order quantum corrections to the field equations. Let the scalar field ϕ_a be in the 3-representation. The

invariant part of the Lagrangian is then:

$$\mathcal{L}^{\text{inv}} = -\tfrac{1}{4}(F_{\mu\nu}^a)^2 - \tfrac{1}{2}(D_\mu\phi_a)^2 - V(\phi) \; ; \qquad V(\phi) = \tfrac{1}{8}\lambda\Big((\phi_a)^2 - F^2\Big)^2 \; . \tag{5.1}$$

Here, D_μ stands for the covariant derivative: $D_\mu\phi_a = \partial_\mu\phi_a + g\epsilon_{abc}A_\mu^b\phi_c$. As in section 2.2, Eq. (2.17), we define shifted fields $\tilde{\phi}_a$ by

$$\phi_a \equiv \tilde{\phi}_a + \begin{pmatrix} 0 \\ 0 \\ F \end{pmatrix} \; ; \qquad V(\tilde{\phi}) = \tfrac{1}{2}\lambda F^2\tilde{\phi}_3^2 + \tfrac{1}{2}\lambda F\tilde{\phi}^2\tilde{\phi}_3 + \tfrac{1}{8}\lambda(\tilde{\phi}^2)^2 \; . \tag{5.2}$$

The shift must also be carried out in the kinetic term for ϕ:

$$D_\mu\phi_a = D_\mu\tilde{\phi}_a + gF\begin{pmatrix} A_\mu^2 \\ -A_\mu^1 \end{pmatrix} \; ; \qquad -\tfrac{1}{2}(D_\mu\phi_a)^2 =$$
$$-\tfrac{1}{2}(D_\mu\tilde{\phi}_a)^2 - gF\Big(A_\mu^2 D_\mu\tilde{\phi}_1 - A_\mu^1 D_\mu\tilde{\phi}_2\Big) - \tfrac{1}{2}g^2F^2\Big(A_\mu^{1\,2} + A_\mu^{2\,2}\Big) \; . \tag{5.3}$$

Defining the complex fields

$$\tilde{\Phi} = \tfrac{1}{\sqrt{2}}(\tilde{\phi}_1 + i\tilde{\phi}_2) \; ; \qquad \mathcal{A}_\mu = \tfrac{1}{\sqrt{2}}(A_\mu^1 + iA_\mu^2) \; ;$$
$$D_\mu\tilde{\Phi} = (\partial_\mu + iA_\mu^3)\tilde{\Phi} - i\mathcal{A}_\mu\tilde{\phi}_3 \; , \tag{5.4}$$

we see that the Lagrangian (5.1) becomes

$$\begin{aligned}
\mathcal{L}^{\text{inv}} &= -\tfrac{1}{4}(F_{\mu\nu}^a)^2 - \tfrac{1}{2}(D_\mu\tilde{\phi}_3)^2 - D_\mu\tilde{\Phi}^* D_\mu\tilde{\Phi} \\
&\quad -\tfrac{1}{2}M_H^2\tilde{\phi}_3^2 - M_V^2\mathcal{A}_\mu^*\mathcal{A}_\mu + M_V\Im(\mathcal{A}_\mu^* D_\mu\tilde{\Phi}) - V^{\text{int}}(\tilde{\phi}) \; , \\
\text{where} \qquad & M_H = \sqrt{\lambda}F \; ; \qquad M_V = gF \; , \tag{5.5}
\end{aligned}$$

and V^{int} is the remainder of the potential term. $\Im$ stands for imaginary part.

Thus, the 'neutral' component of the scalar field, the Higgs particle, gets a mass M_H (see Eq. 5.2) and the 'charged' components of the vector field receive a mass term with mass M_V. The mechanism that removes (some of) the Goldstone bosons and generates mass for the vector particles, is called the Brout-Englert-Higgs (BEH) mechanism.[10][11] In every respect, the neutral, massless component of the vector field behaves like an electromagnetic vector potential, and the complex vector particle is electrically charged.

5.2. Fixing the gauge

If one would try to use the rules of Subsection 4.4 to derive the Feynman rules directly from $\mathcal{L}^{\text{inv}}$, one would find that the matrix $\hat{M}$ describing the bilinear part of the Lagrangian

has no inverse. This is because the gauge must first be fixed. Choosing $\partial_\mu A^a_\mu(x) = 0$ has the advantage that the somewhat awkward term $\Im(\mathcal{A}^*_\mu \partial_\mu \tilde{\Phi})$ can be put equal to zero by partial integration. The vector propagator (in momentum space) is then easily computed to be

$$P^{ab}_{\mu\nu}(k) = \frac{\delta_{\mu\nu} - k_\mu k_\mu/(k^2 - i\varepsilon)}{k^2 + m^2_{(a)} - i\varepsilon} \, \delta_{ab} \, , \tag{5.6}$$

where $m_{(a)} = M_V$ for the charged vector field and 0 for the neutral one.

This indeed appears to describe a vector particle with mass $m_{(a)}$ and an additional transversality constraint. One can do something smarter, though. If, in the gauge-fixed lagrangian (4.36), we choose

$$C^3 = \partial_\mu A^3_\mu \; ; \qquad C^1 = \partial_\mu A^1_\mu - M_V \tilde{\phi}_2 \; ; \qquad C^2 = \partial_\mu A^2_\mu + M_V \tilde{\phi}_1 \; , \tag{5.7}$$

then we find that the scalar-vector mixing terms cancel out, but now also the $(\partial_\mu A_\mu)^2$ term cancels out, so that the vector propagator looses its $k_\mu k_\nu$ term. The vector propagator is then

$$P^{ab}_{\mu\nu}(k) = \frac{\delta_{\mu\nu}\delta_{ab}}{k^2 + m^2_{(a)} - i\varepsilon} \, , \tag{5.8}$$

and the charged scalar ghost gets a mass M_V. The physical field $\tilde{\phi}_3$ is unaffected.

It is instructive to compute the Faddeev-Popov ghost Lagrangian in this gauge. One easily finds it to be

$$\mathcal{L}^{\text{ghost}} = \overline{\eta}^a \partial^2 \eta^a - M^2_V(\overline{\eta}^1 \eta^1 + \overline{\eta}^2 \eta^2) + \text{ interaction terms} \, . \tag{5.9}$$

As will be confirmed by more explicit calculations, the theory has physical, charged vector particles with masses M_V, a neutral (massless) photon and a neutral scalar particle with mass M_H. The latter is called the Higgs particle of this theory. All other fields in the Lagrangian describe ghost fields. Apparently, in the gauge described above, all 'unphysical' charged particles, the ghosts, the timelike components of the vector fields, as well as the Goldstone bosons, have the same mass M_V. The unphysical neutral particles all have mass zero.

One concludes that the symmetry pattern of this example is as follows: the local gauge group, $SO(3)$, is broken by the Brout-Englert-Higgs mechanism into its subgroup $SO(2)$ (the rotations about a fixed axis, formed by the vacuum value of ϕ_a), or equivalently, $U(1)$. Therefore two of the three vector bosons obtain a mass, while one massless $U(1)$ photon remains. At the same time, two of the three scalars turn into ghosts, the third into a Higgs particle.

37

The Brout-Englert-Higgs mechanism does not alter the total number of independent physical states in the particle spectrum. In our example, two of the three scalar particles disappeared, but the two massive spin-1 particles now each have three spin helicities, whereas the massless photons only had two.

5.3. Coupling to other fields

The shift (5.2) in the definition of the fields, gives all interactions an asymmetric appearance. This is why, in the literature, one talks of "spontaneous breaking of the local symmetry". Actually, this is something of a misnomer. In the case of a *global* symmetry, spontaneous breakdown means that the vacuum state is degenerate. After a global symmetry transformation, the vacuum state is transformed into a physically inequivalent vacuum state, which is not realized in the system. The existence of a massless Goldstone boson testifies to that. In the case of a *local* symmetry, nothing of the sort happens. There is only one vacuum state, and it is invariant under the local symmetry, always. This is why the Goldstone boson became unphysical. In fact, *all* physical states are formally invariant under local gauge transformations. Apparent exceptions to this rule are, of course, the charged particles in QED, but this is because we usually wish to ignore their interactions with the vector potential at infinity. In reality a full discussion of charged particles is obscured by their long-range interactions.

In view of all of this, it is better not to say that a local symmetry is spontaneously broken, but, rather, to talk of the *Brout-Englert-Higgs mechanism* [10][11], which is the phenomenon that the spectrum of physical particles do not form a representation of the local symmetry group. The local symmetry can only be recognized by shifting the scalar fields back to their symmetric notation, the original fields ϕ. Local symmetry must not be regarded as a property of the physical states, but rather as a property of our way of describing the physical states.

If, however, we perform a perturbation expansion for small values of the gauge field coupling, we find that at vanishing gauge coupling a local symmetry *is* spontaneously broken. Therefore, it is still quite useful to characterize our perturbative description by listing the gauge groups and the subgroups into which they are broken.

Now, let us assume that there are other fields present, such as the Dirac fermions, ψ_i. In the symmetric notation, they must form a representation of the local gauge group. So, we have

$$\mathcal{L}^{\text{Dirac}} = -\overline{\psi}^i(\gamma_\mu D_\mu + m_{(i)}\psi_i - \overline{\psi}^i g_Y t^a_{ij}\phi_a\psi_j) \ , \tag{5.10}$$

where D_μ is the appropriate covariant derivative, containing those matrices T^a that are appropriate for the given representation (see 4.11 and 4.12), and g_Y stands for one or more Yukawa coupling parameters. The mass terms $m_{(i)}$ and coupling coefficients t^a_{ij}

are invariant tensors of the gauge group (masses are only allowed if the fermions are not chiral, see the discussion following Eq. (4.22)).

Here, again, we started with the more transparent symmetric fields ϕ_a, but the physical fields $\tilde{\phi}$ are obtained by the shift $\phi_a = F_a + \tilde{\phi}_a$. Thus, the lowest order bilinear part of the Dirac Lagrangian becomes

$$\mathcal{L}^{\text{Dirac}} \rightarrow -\overline{\psi}^i \Big((\gamma_\mu \partial_\mu + m_{(i)}) \delta_{ij} + g_Y t^a_{ij} F_a \Big) \psi_j \ , \tag{5.11}$$

In particular, if the symmetry acts distinctly on the chiral parts of the fermion fields, the mass term $m_{(i)}$ is forbidden, but the less symmetric second term may generate masses and in any case mass differences for the fermions. Thus, not only do the vector and scalar particles no longer form representations of the original local gauge group, but neither do the fermions.

5.4.   The Standard Model

What is presently called the 'Standard Model' is just an example of a Higgs theory. The gauge group is $SU(3) \times SU(2) \times U(1)$. This means that the set of vector fields falls apart into three groups: 8 associated to $SU(3)$, then 3 for $SU(2)$, and finally one for $U(1)$. The scalar fields ϕ_i form one two-dimensional, complex representation of two of the three groups: it is a doublet under $SU(2)$ and rotates as a particle with charge $\frac{1}{2}$ under $U(1)$.

Representing the Higgs scalar in terms of four real field components, the Brout-Englert-Higgs mechanism is found to remove three of them, leaving only one neutral, physical Higgs particle. $SU(2) \times U(1)$ is broken into a diagonal subgroup $U(1)$. Three of the four gauge fields gain a mass. The one surviving photon field is obtained after re-diagonalizing the vector fields; it is a linear composition of the original $U(1)$ field and one of the three components of the $SU(2)$ gauge fields.

The $SU(3)$ group is not affected by the Brout-Englert-Higgs mechanism, so one would expect all 'physical' particles to come in representations of $SU(3)$. What happens instead is further explained in section 11: only gauge-invariant combinations of fields are observable as particles in our detectors.

The fermions in the Standard Model form three 'families'. In each family, we see the same pattern. The left handed fields, ψ_L, all form doublets under $SU(2)$, and a combination of a triplet ('quarks') and singlets ('leptons') under $SU(3)$. The right handed components, ψ_R, form the same representations under $SU(3)$, but form a pair of two singlets under $SU(2)$; so they do not couple to the $SU(2)$ vector fields. The $U(1)$ charges of the left-handed $SU(2)$ doublets are $-\frac{1}{2}$ for the leptons and $\frac{1}{6}$ for the quarks; the U(1) charges of the right-handed singlets are -1 and 0 (for the leptons) or $-\frac{1}{3}$ and $\frac{2}{3}$ (for the quarks).

The Standard model owes its structure to the various possible Yukawa interaction terms with the Higgs scalars. They are all of the form $\overline{\psi}\,\phi\,\psi$, and invariant under the entire gauge group, but since there are three families of fermions, each having left and right handed chiral components, there are still a fairly large number of such terms, each of which describes an interaction strength whose value is not dictated by the principles of our theory.[12]

6. Unitarity

As we saw in subsection 4.4, the Feynman rules unambiguously follow from the expression one has for the Lagrangian of the theory. More precisely, what was derived there was the set of rules for the vacuum-to-vacuum amplitude in the presence of possible source insertions $J_i(x)$, including anticommuting sources η_i, $\overline{\eta}_i$. The overall multiplicative constant C in our Gaussian integrals such as (2.46) is completely fixed by the demand that, in the absence of sources, the vacuum-to-vacuum amplitude should be 1. By construction then, the resulting scattering matrix should turn out to be unitary.

In practice, however, things are not quite that simple. In actual calculations, one often encounters divergent, hence meaningless expressions. This happens when one makes the transition to the continuum limit too soon — remember that we insisted that space and time are first kept discrete. Unitarity of the S-matrix turns out to be a sensitive criterion to check whether we are performing the continuum limit correctly. It was one of our primary demands when we initiated the program of constructing workable models for relativistic, quantized particles. Another demand, the validity of dispersion relations, can be handled the same way as unitarity; these two concepts will be shown to be closely related. The formalism described below is based on work by Cutkosky and others, but was greatly simplified by Veltman.[13]

Parts of this section are fairly technical and could be skipped at first reading.

6.1. The largest time equation

Let us start with the elementary Feynman propagator, $(k^2 + m^2 - i\varepsilon)^{-1}$, and its Fourier transform back to configuration space (omitting for simplicity a factor $(2\pi)^4$):

$$\Delta^F(x) = -i \int \mathrm{d}^4 k \, \frac{e^{ikx}}{k^2 + m^2 - i\varepsilon} \,, \qquad x = x^{(1)} - x^{(2)} \,. \tag{6.1}$$

In addition, we define the *on-shell propagators*

$$\Delta^\pm(x) = 2\pi \int \mathrm{d}^4 k \, e^{ikx} \delta(k^2 + m^2)\theta(\pm k^0) \,; \qquad k\,x = \vec{k} \cdot \vec{x} - k^0 x^0 \,, \tag{6.2}$$

and θ is the Heaviside step function, $\theta(x) = 1$ for $x \geq 0$ and $= 0$ otherwise. The integrals are over Minkowski variables $\vec{k}$, k^0. These operators propagate particles on mass shell with the given sign of the energy from $x^{(2)}$ to $x^{(1)}$, or back with the opposite sign. We have

$$\Delta^+(x) = (\Delta^-(x))^* \; ; \qquad \Delta^+(x) = \Delta^-(-x) \; . \tag{6.3}$$

Our starting point is the decomposition of the propagator into forward and backward parts:

$$\Delta^F(x) = \theta(x^0)\Delta^+(x) + \theta(-x^0)\Delta^-(x) \; . \tag{6.4}$$

Obviously:

$$\Delta^{F*}(x) = \theta(x^0)\Delta^-(x) + \theta(-x^0)\Delta^+(x) \; . \tag{6.5}$$

One easily proves this by deforming the contour integration in the complex k^0 plane.

Consider now a Feynman diagram with n vertices, where lines are attached with a given topological structure, which will be kept fixed. The external lines are assumed to be 'amputated': there are no propagators attached to them. The Feynman rules are applied as described in Subsections 2.5 and 4.4. The diagram is then part of our calculation of an S-matrix element. We consider the diagram in momentum representation *and* in the coordinate representation. The expression we get in coordinate representation is called $F(x^{(1)}, x^{(2)}, \cdots, x^{(n)})$.

Next, we introduce an expression associated to the same diagram, but where some of the vertices are underlined: $F(x^{(1)}, x^{(2)}, \cdots, \underline{x}^{(i)}, \cdots, \underline{x}^{(j)}, \cdots, x^{(n)})$, where $x^{(i)}$ refer to the coordinates that must be integrated over when one elaborates the Feynman rules. The rules for computing this new amplitude are as follows:

i) A propagator $\Delta^F(x^{(i)} - x^{(j)})$ is used if neither $x^{(i)}$ nor $x^{(j)}$ are underlined.

ii) A propagator $\Delta^+(\underline{x}^{(i)} - x^{(j)})$ is used if $x^{(i)}$ but not $x^{(j)}$ is underlined.

iii) A propagator $\Delta^-(x^{(i)} - \underline{x}^{(j)})$ is used if $x^{(j)}$ but not $x^{(i)}$ is underlined.

iv) A propagator $\Delta^{F*}(\underline{x}^{(i)} - \underline{x}^{(j)})$ is used if both $x^{(i)}$ and $x^{(j)}$ are underlined.

v) A minus sign is added for every underlined vertex.

In all other respects, the rules for the calculation of the amplitude are unchanged.

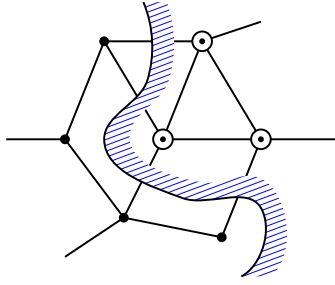One now derives the *largest time equation*:

Figure 3: Diagram with underlined vertices, which are indicated by little circles

Let $x^{(k)}$ be the coordinate with the largest time:

$$x^{(k)0} \geq x^{(i)0} \ , \quad \forall i \ .$$

Then,

$$F(x^{(1)}, x^{(2)}, \cdots, x^{(k)}, \cdots, \underline{x}^{(n)}) = -F(x^{(1)}, x^{(2)}, \cdots, \underline{x}^{(k)}, \cdots, \underline{x}^{(n)}) \ , \qquad (6.6)$$

where in both terms the points other than $x^{(k)}$ are underlined or not in identical ways.

One easily proves this using Eqs. (6.4) and (6.5). One consequence of this theorem is

$$\sum_{\text{all } 2^n \text{ possible underlinings}} F(\{x^{(i)}\}) = 0 \ . \qquad (6.7)$$

We now show that these are the diagrams contributing to the unitarity equation, or 'optical theorem':

$$\sum_n S|n\rangle\langle n|S^\dagger = \mathbb{I} \ . \qquad (6.8)$$

The diagrams for the matrix S are as described earlier. The diagrams for $S^\dagger$ contain the complex conjugates of the propagators. Since also the vertices in the functional integral are all multiplied by i, they must all change sign in $S^\dagger$. Also the momenta k in e^{ikx} switch sign. In short, the diagrams needed for the computation of $S^\dagger$ indeed are the underlined Green functions. Note that, in momentum space, the largest time equation (6.6) cannot be applied to individual vertices, since, while being integrated over, the vertex with largest time switches position. However, the summed equation (6.7) is valid. The identity $\mathbb{I}$ on the r.h.s. of Eq. (6.8) comes from the one structure that survives: the diagram with no vertices at all.

42

We observe that unitarity may follow if we add all possible ways in which a diagram with given topology can be cut in two, as depicted in Fig. 3. The shaded line separates S from $S^\dagger$.

The lines joining S to $S^\dagger$ represent the intermediate states $|n\rangle$ in Eq. (6.8). They are on mass shell and have positive energies, which is why we need the factors $\delta(k^2+m^2)\theta(k^0)$ there. If a propagator is equipped with some extra coefficients R_{ij}:

$$P_{ij}(k) = \frac{-i\,R_{ij}(k)}{k^2 + m^2 - i\varepsilon} \ , \tag{6.9}$$

then we can still use the same decomposition (6.4), provided R_{ij} is *local*: it must be a finite polynomial in k. Writing

$$R_{ij} = \sum_k f_i(k) f_j^*(k) \ , \tag{6.10}$$

we can absorb the factors $f_i(k)$ into the definition of S, *provided that all eigenvalues of R_{ij} are non-negative.* Indeed, kinetic terms in the Lagrangian must all have the same sign.

Note that we are *not* allowed to replace the terms in the Lagrangian by their complex conjugates. This implies that, for the unitarity proof, it is mandatory that the Lagrange density is a real function of the fields.

An important feature of these equations is the theta functions for k^0. They guarantee that the intermediate states contribute only if their total energy does not exceed the energy available in the given channel.

6.2. Dressed propagators

In the previous subsection, not all diagrams that contribute to $S\,S^\dagger$ have yet been handled correctly. There is a complication when self-energy diagrams occur. If one of the lines at both sides of a self-energy blob is replaced by $\Delta^\pm$, then the other propagator Δ^F places a pole on top of that Dirac delta. In this case, we have to use a more sophisticated prescription. To see what happens, we must first sum the geometric series of propagator insertions, see Fig. 4(a). We obtain what is called the dressed propagator. In momentum space, let us write the contribution of a single blob in Fig. 4(a) as $-i\delta M(k)$. It represents the summed contribution of all *irreducible* diagrams, which are the diagrams with two external lines that cannot fall apart if one cuts one internal line. We need its real and imaginary parts: $\delta M(k) \equiv \delta m^2(k) - i\Gamma(k)$. Write the full propagator as

$$\begin{aligned}
P^{\mathrm{dr}}(k) &= P^0(k) - P^0(k)i\delta M(k)P^0(k) + \cdots \\
&= P^0(k)\sum_{n=0}^{\infty}\Big(-i\delta M(k)P^0(k)\Big)^n \ = \ \frac{P^0(k)}{1 + i\delta M(k)P^0(k)} \ ; \tag{6.11}
\end{aligned}$$

43

$$\text{if} \quad P^0(k) \;=\; -i(M(k) - i\varepsilon)^{-1}$$

$$\text{then} \qquad P^{\mathrm{dr}}(k) = -i(M(k) + \delta M(k) - i\varepsilon)^{-1} \;, \tag{6.12}$$

where $P^0(k)$ is the unperturbed ('bare') propagator.

If we define the real part of the dressed propagator (in momentum space) to be

$$\Re(P^{\mathrm{dr}}(k)) = \frac{\Gamma(k)}{(k^2 + M + \delta m^2)^2 + \Gamma^2} = \pi\varrho(-k^2) \;, \tag{6.13}$$

then, by contour integration,

$$P^{\mathrm{dr}}(k) = \int_0^\infty \mathrm{d}m^2 \frac{\varrho(m^2)}{k^2 + m^2 - i\varepsilon} \;; \tag{6.14}$$

we call this the Källen-Lehmann representation of the propagator. Later, it will be assured that $\varrho(m^2) = 0$ if $m^2 < 0$.
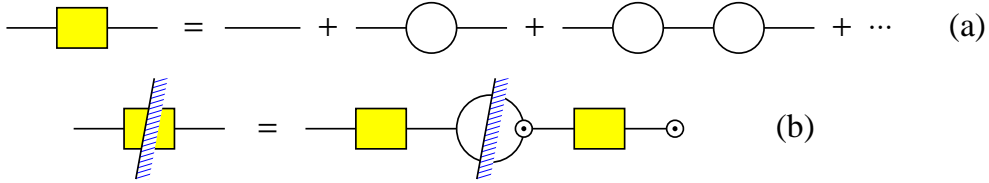

Figure 4: (a) The dressed propagator as a geometric series;
(b) Cutting the dressed propagator

The best strategy now is to apply a largest time equation to the entire dressed propagator. Write as for Eqs. (6.4) and (6.5),

$$P^{\mathrm{dr}}(x) \;=\; \theta(x^0)\Delta^+_{\mathrm{dr}}(x) + \theta(-x^0)\Delta^-_{\mathrm{dr}}(x) \;;$$
$$P^{\mathrm{dr}}(x)^* \;=\; \theta(x^0)\Delta^-_{\mathrm{dr}}(x) + \theta(-x^0)\Delta^+_{\mathrm{dr}}(x) \;. \tag{6.15}$$

Then,

$$\Delta^{\pm}_{\mathrm{dr}}(k) = 2\pi \int \mathrm{d}^4k \, e^{ikx} \varrho(-k^2)\theta(\pm k^0) \;. \tag{6.16}$$

The imaginary part $\Gamma(k)$ of the irreducible diagrams can itself again be found by applying the cutting rules. Writing $S = \mathbb{1} + iT$, we find that unitarity for all non-trivial diagrams corresponds to $i(T - T^\dagger) + T T^\dagger = 0$, and the diagrams for $T T^\dagger$ are depicted in Fig. 4b. They are exactly the diagrams needed for unitarity of the entire scattering matrix involving a single virtual particle in the channel of two external ones.

One observes that the function $\varrho(-k^2)$ must be non-negative, and only nonvanishing for timelike k. The latter is guaranteed by the theta functions in k^0. Only the delta peaks in ϱ are associated to *stable* particles that occur in the initial and final states of the scattering matrix. Resonances with finite widths contribute to the unitarity of the scattering matrix via their stable decay products.

44

6.3.   Wave functions for in- and out-going particles

Many technical details would require too much space for a full discussion, so we have to keep this sketchy. In case we are dealing with vector or spinor particles, the residues R_{ij} of the propagators represent the summed absolute squares of the particle wave functions. We have seen in Eq. (6.9) how this comes about. If, for example, a vector particle is described by a propagator

$$P_{\mu\nu} = -i\frac{\delta_{\mu\nu} + k_\mu k_\nu/M^2}{k^2 + M^2 - i\varepsilon} \ , \tag{6.17}$$

then we see that, first of all, the numerator is a polynomial in k, as was required, and, if we go on mass shell, $k^2 = -M^2$, then we see that the field component proportional to k_μ is projected out. In particular, if we put $k = (0, 0, 0, iM)$, then $R_{ij} = \delta_{ij}$ and its timelike components disappear, so indeed there are three independent states for the particle described.

For the fermions, the bare propagator is

$$P^{\text{Dirac}} = -i\frac{m - i\gamma k}{k^2 + m^2 - i\varepsilon} \ . \tag{6.18}$$

Before relating this to the renormalization of the wave functions, we must note that all γ^μ are hermitean, while k_i are real and k_4 is imaginary. We observe that the Feynman rules for $S^\dagger$ are like those of S, but with γ^4 replaced by $-\gamma^4$. Next, the arrows in the propagators must be reversed. This leads to an extra minus sign for every vector k_μ, while γ^μ are replaced by $\gamma^{\mu\dagger}$. All together, one requires that $\gamma^i \to -\gamma^i$ while γ^4 remains unchanged. This amounts to the replacement $\gamma^\mu \to \gamma^4\gamma^\mu\gamma^4$. One concludes that the rules for $S^\dagger$ are like those for S if all fermion lines enter or leave the diagram with an extra factor γ^4. This means that the wave functions for external fermions in a diagram are to be normalized as

$$(m - i\gamma_i k^i + \gamma^4 k^0)\gamma^4 = \sum_{i=1}^{2} |\psi_i(k)\rangle\langle\psi_i(k)| \ , \qquad (k^0 > 0) \ ; \tag{6.19}$$

while for the anti-fermions, we must demand

$$\gamma^4(m - i\gamma_i k^i - \gamma^4 k^0) = -\sum_{i=3}^{4} |\psi_i(k)\rangle\langle\psi_i(k)| \ , \qquad (k^0 > 0) \ . \tag{6.20}$$

The minus sign is necessary because the operator in (6.20) has two negative eigenvalues. One concludes that unitarity requires spin-$\frac{1}{2}$ particles to carry one extra minus sign for each closed loop of these particles. This leads to the necessity of Fermi-Dirac statistics. Again, it is important that none of the higher order corrections ever affect the signs of the

eigenvalues for these projection operators, since these can never be accommodated for by a renormalization of the particle wave function.

The conclusion of this section, that the scattering matrix is unitary in the space of physical particle states, should not come as a surprise because our theory has been constructed to be that way. Yet it is important that we see here in what way the Feynman diagrams intertwine to produce unitarity explicitly.

We also see that unitarity is much more difficult to control when we have ghosts due to the gauge fixing procedure. Our vector particles then have propagators where Eq. (6.17) is replaced by expressions such as

$$P^{\text{ren}}_{\mu\nu} = \frac{-i \, g_{\mu\nu}}{k^2 + M^2 - i\varepsilon} \; . \tag{6.21}$$

We write here $g_{\mu\nu}$ rather than $\delta_{\mu\nu}$ in order to emphasize that our arguments are applied in Minkowski space, where clearly the time components 'carry the wrong sign'. The field components associated to that would correspond to particles that contribute negatively to the scattering probabilities. To correct this, one would have to replace $|n\rangle\langle n|$ by $-|n\rangle\langle n|$, which cannot be achieved by renormalizing the states $|n\rangle$. Here, we use the BRST relations to show that all unphysical states can be transformed away. In practice, we use the fact that the scattering matrix does not depend on the choice of the gauge fixing function $C^a(x)$, so we choose it such that all ghost particles have a mass exceeding some critical value Λ. *In the intermediate states, their projector operators $\Delta^{\pm}(k)$ then only contribute if the total energy in the given channel exceeds Λ.* This then means that there are no ghosts in the intermediate states, so the scattering matrix is unitary in the space of physical particles only — an absolutely essential step in the argument that these theories are internally consistent. The required gauge fixing functions $C^a(x)$ are not difficult to construct, but their existence is only needed to complete this formal argument. They are rather cumbersome to use in practical calculations.

6.4. Dispersion relations

The largest time equation can also be employed to derive very important *dispersion relations* for the diagrams. These imply that any diagram D can be regarded as a combination of two sets of diagrams D_i and $D_i^\dagger$:

$$D = \sum_i \int_0^\infty \frac{\mathrm{d}k^0}{-k^0 - i\varepsilon} \, D_i(k^0) \, D_i^\dagger(k^0) \; . \tag{6.22}$$

Here, $D_i(k^0)$ and $D_i^\dagger(k^0)$ stand for amplitudes depending on various external momenta k, where one of the timelike components, k^0, is integrated over. This, one derives by singling out two points, $x^{(1)}$ and $x^{(2)}$ in a diagram, and time-ordering them. The details

of the derivation go beyond the scope of this paper (although they are not more difficult than the previous derivations in this section). Eq. (6.22) can be used to express diagrams with closed loops in terms of diagrams with fewer closed loops, and discuss the subtraction procedures needed for renormalization.

7. Renormalization

For a proper discussion of the renormalization concept, we must emphasize what our starting point was: first, replace the continuum of space by a dense lattice of points, and only at the very end of all calculations do we make an attempt to take the continuum limit. The path integral procedure, illuminated in subsection 2.4, implies that time, also, can be replaced by a lattice. In Fourier space, the space-time lattice leads to finite domains for the values of energies and momenta (the Brillouin zones), so that all ultra-violet divergences disappear. If we also wish to ensure the absence of infra-red divergences, we must replace the infinite volume of space and time by a finite box. This is often required if complications arise due to divergent contributions of soft virtual particles, typically photons. Nasty infra-red divergences occur in theories with confinement, to be discussed in section 11.

The instruments that we shall use for the ultra-violet divergences of a theory are as follows. We assume that all freely adjustable physical constants of the theory, referred to as the 'bare' parameters, such as the 'bare' mass and charge of a particle, should be carefully tuned to agree with observation, but the tuning process may depend critically on the mesh size a of the space-time lattice. Thus, while we vary a, we allow all bare parameters, λ say, in the theories to depend on a, often tending either to infinity or to zero as $a \to 0$. If this procedure is combined with a perturbation expansion, say in terms of a small coupling g, we expect to find that observable features depend minimally on a provided that the bare couplings $g(a)$ remain small in the limit $a \downarrow 0$.

This will be an important condition for our theories to make sense at all. How do we know whether $g(a)$ tends to zero or not? The simplest thing to look at, is the *dimensionality* of g. All parameters of a field theory have a dimension of a length to some power. These dimensions usually depend on the dimensionality n of space-time. The rules to compute them are easy to obtain:

- An action $S = \int \mathrm{d}^n x \mathcal{L}(x)$ is dimensionless;

- The dimension of a Lagrange density $\mathcal{L}$ is therefore $(\text{length})^{-n} = m^n$, where m is a mass.

- The dimension of the fields can be read off from the kinetic terms in the Lagrangian, because they contain no further parameters. A scalar field ϕ has dimension $m^{(n-2)/2}$,

a fermionic field ψ has dimension $m^{(n-1)/2}$.

- A gauge coupling constant g has dimension $m^{(4-n)/2}$ and the coupling parameter λ in an interaction term of the form $\lambda\phi^k$ has dimension $m^{n+k-nk/2}$,

and so on.

> *A theory is called power-counting renormalizable, if all expansion parameters have mass-dimension positive or zero.*

This is why, in 4 space-time dimensions, we cannot accept higher than quartic interactions among scalars. In practice, in 4 space-time dimensions, most expansion parameters have dimension zero. In Section 9, we will see that dimensionless coupling parameters nevertheless depend on the size of a, but only logarithmically:

$$\lambda(a) \approx \lambda_0 + C\lambda_0^2 \log(a) \ + \text{higher orders.} \tag{7.1}$$

Regardless of whether this tends to zero or to infinity in the continuum limit, one finds that, in the continuum theory, the perturbative corrections to the bare parameters λ diverge. This is nothing to be alarmed about. However, if λ itself is also a small parameter in terms of which we wish to perform a perturbation expansion, then clearly trouble is to be expected if its bare value tends to infinity. Indeed, we shall argue that, in general, such theories are inconsistent.

There are two very important remarks to be made:

— Theories can be constructed where all couplings really tend to zero in the continuum limit. These theories are called *asymptotically free* (Section 9), and they allow for accurate approximations in the ultra-violet. It is generally believed that such theories can be defined in a completely unambiguous fashion through their perturbation expansions in the ultra-violet; in any case, they allow for very accurate calculation of all their physical properties. QCD is the prime example.

— If a theory is not asymptotically free, but has only small coupling parameters, the perturbation expansion formally diverges, and the continuum limit formally does not exist. But the first N terms of perturbation expansion do make sense, where $N = \mathcal{O}(1/g)$. This means that uncontrollable margins of error are exponentially small, of order $e^{-C/g}$ or e^{-C/g^2}, which in practice is much smaller than other uncertainties in the theory, so they are of hardly any practical consequence. Thus, in such a case, our theory *does* have intrinsic inaccuracies, but these are *exponentially suppressed*. In practice, such theories are still highly valuable. The Standard Model is an example.

A useful approach is to substitute all *numbers* in a theory by formal *series expansions*, where the expansion parameter, a factor common to all coupling parameters of the theory, is formally kept infinitesimal. In that case, all perturbation coefficients are uniquely defined, though one has little direct knowledge concerning the convergence or divergence of the expansions.

In both the cases mentioned above, our theories are defined from their perturbation expansion; clearly, the perturbation expansion is not only a convenient device for calculations, it is an essential ingredient in our theories. Let us therefore study how renormalization works, order-by-order in perturbation theory.

In a connected diagram, let the number of external lines be E, the number of propagators be P, and let V_n be the number of vertices with n prongs. By drawing two dots on each propagator and one on each external line, one finds that the number of dots is

$$2P + E \;=\; \sum_n nV_n = 3V_3 + 4V_4. \tag{7.2}$$

For tree diagrams (simply connected diagrams), one finds by induction, with V the number of vertices, $V = \sum_n V_n$,

$$V = P + 1 \;. \tag{7.3}$$

A diagram with L closed loops (an L-fold connected diagram) turns into a tree by cutting away L propagators. Therefore, one has

$$P = V - 1 + L \;. \tag{7.4}$$

Combining Eqs. (7.2) and (7.4), one has

$$E + 2L - 2 \;=\; \sum_n (n-2)V_n \;=\; V_3 + 2V_4 \;. \tag{7.5}$$

Consequently, if every 3-vertex comes with a factor g and every 4-vertex with a factor λ, and if a diagram with a given number E of external lines, behaves as $g^{2n}\lambda^k$, it must have $L = n + k + 1 - \frac{1}{2}E$ closed loops. Perturbation expansion is therefore often regarded as an expansion in terms of the number of closed loops.

7.1. Regularization schemes

In a tree diagram, in momentum space, no integrations are needed to be done — the momentum flowing through every propagator is fixed by the momenta of the in- and out-going particles. But if there are L loops, one has to perform $4L$ integrations in momentum space. It is these integrations that often tend to diverge at large momenta.

Of course, these divergences are stopped if momentum space is cut off, as is the case in a finite lattice. However, since our lattice is not Lorentz-invariant and may lack other symmetries such as gauge-invariance, it is useful to find *other ways to modify our theory so that UV divergences disappear*. This is called 'regularization'. We give two examples.

7.1.1. Pauli-Villars regularization

Assume that a propagator of the form shown is replaced as follows:

$$\frac{A(k)}{k^2 + m^2 - i\varepsilon} \quad \rightarrow \quad \sum_i e_i \frac{A(k)}{k^2 + \Lambda_i^2 - i\varepsilon} \ ; \quad \sum_i e_i = 0 \ , \ \sum_i e_i \Lambda_i^2 = 0 \ . \tag{7.6}$$

If we take $e_1 = 1$, $\Lambda_1 = m$, while all other Λ_i tend to $+\infty$, we get back the original propagator. With finite Λ_i, however, we can make all momentum integrations converge at infinity. Our theory is then finite. This is (a somewhat simplified version of) Pauli-Villars regularization.

However, the new propagators cannot describe ordinary particles. The ones with $e_i < 0$ contribute to the unitarity relation with the wrong sign! On the other hand, the $i\varepsilon$ prescription is as usual, so that these particles do carry positive energy. In any channel where the total energy is less than Λ_i, the 'Pauli-Villars ghosts' do not contribute to the unitarity relation at all. So, in a theory where we put a limit to the total energy considered, Pauli-Villars regularization is physically acceptable. In practice, we will try to send all ghost masses Λ_i to infinity.

7.1.2. Dimensional regularization

Dimensional regularization[9] consists of formally performing all loop integrations in $4 - \varepsilon$ dimensions, where ε may be any (possibly complex) number. As long as ε is irrational, all integrations can be replaced by finite expressions following an unambiguous prescription, to be explained below. If $\varepsilon = 0$, one can also subtract the integrals, but the prescription is then often not unambiguous, so that *anomalies* might arise. This is why dimensional regularization will be particularly important whenever the emergence of anomalies is a problem one wishes to understand and control.

It is important to realize that also when $\varepsilon \neq 0$, integrals may be divergent, but that, for irrational ε, unambiguous subtractions may be made. This needs to be explained, but first, one needs to define what it *means* to have non-integral dimensions. Such a definition is only well understood within the frame of the perturbation-, or loop-, expansion. It is obvious from the construction of the theory, that a diagram with L loops involves integrals over integrands that are purely rational functions in $L(4 - \varepsilon)$ variables. Observing that the external lines (of which we have only a finite number) span some N dimensional space, we now employ the fact that the remaining variables occur in a rotationally symmetric

fashion. There, we write the formula for the ℓ-dimensional (Euclidean) sphere of radius r as

$$\int \mathrm{d}^\ell k \, \delta(k^2 - r^2) = \frac{\pi^{\ell/2}}{\Gamma(\ell/2)} r^{\ell-2} \ . \tag{7.7}$$

Here, Γ stands for the Euler gamma function, $\Gamma(z) = (z - 1)!$ for integral z.

It is at this point where we can decide that this expression *defines* the integral for any, possibly complex, value for ℓ. It converges towards the usual values whenever ℓ happens to be a positive integer. After having used Eq. (7.7), one ends up with an integral over s variables k_μ of a function $f(k)$, where s is an integer, but $f(k)$ contains ε-dependent powers of polynomials in k.

Convergence or divergence of an integral can be read off from simple power counting arguments, and, at first sight, one sees hardly any improvement when ε is close to zero. However, what is achieved is that infra-red divergences ($k_\mu \to 0$) are separated from the ultra-violet divergences ($k_\mu \to \infty$), and this allows us to define the "finite parts" of the integrals unambiguously:

- All integrals $\int \mathrm{d}^s k \, f(k)$ are replaced by functionals $I(\{f(k)\})$ that obey the same combinatorial rules as ordinary integrals:

$$\begin{aligned} I(\alpha f_1 + \beta f_2) &= \alpha I(f_1) + \beta I(f_2) \ , \\ I(\{f(k+q)\}) &= I(\{f(k)\}) \ , \end{aligned} \tag{7.8}$$

- $I(f) = \int \mathrm{d}^s k \, f(k)$ if this converges.

- $I(f) = 0$ if $f(k) = (k^2)^p$ when $2p + s$ is *not an integer*.

This latter condition is usually fulfilled, if we started with ε not integer.

These rules are sufficient to replace any integral one encounters in a Feynman diagram by some finite expression. Note, however, that complications arise if one wants to use these rules when $2p + s$ is an integer, particularly when it is zero. In that case, the expression diverges in the ultra-violet *and* in the infra-red, so, in this case, it cannot be used to remove all divergences — it can only replace one by another. Consequently, our finite expressions tend to infinity as $\varepsilon \to 0$.

It is important to verify that dimensional regularization fully respects unitarity and the dispersion relations discussed above. Therefore, the 'dimensionally regularized' diagrams correspond to solutions of the dispersion relations and the unitarity relations, providing some 'natural' subtraction.

7.1.3. Equivalence of regularization schemes

The subtractions provided by the various regularization schemes discussed above, in general, are not the same. At any given order, they do all obey the same dispersion relations of the form (6.22). If we ask, which amplitudes can be added to one scheme to reproduce the other, or, what is the amplitude of the difference between the two schemes (after having eliminated these differences at the order where the subdiagrams $D_i(k^0)$ had been computed), we find the following. This difference must be a Lorentz-covariant expression; and it can only come from the dimensionally regularized contributions of the unphysical Pauli-Villars ghosts in Eq. (6.22). Because of their large masses, only very large values of k^0 in this equation contribute. The p^0-dependence then must reduce to being a polynomial one (p being the momenta of the fixed external lines), and because of Lorentz-invariance, the expression must be polynomial in all components of p_μ. This is exactly what can be achieved by putting a *counter term inside the bare Lagrangian of the theory*. This way, one derives that the various regulators differ from one another by different effective couplings in the bare Lagrangian.

It is then a question of taste which regulator one prefers. Since dimensional regularization often completely respects local gauge-invariance[8], and also because it turned out to be very convenient and efficient in practice, one often prefers that. It should always be kept in mind, however, that dimensional regularization is something of a mathematical trick, and the physical expressions only make sense in the limit $\varepsilon \to 0$.

7.2. Renormalization of gauge theories

Using the results from the previous Sections, we decide to treat quantum field theories in general, and gauge theories in particular, as follows: first, we *regularize* the theory, by using a 'lattice cut-off', or a Pauli-Villars cut-off, or by turning towards $n = 4 - \varepsilon$ dimensions. All these procedures are characterized by a small parameter, such as ε, such that the physical theory is formally obtained in the limit $\varepsilon \to 0$. These procedures are all equivalent, in the sense that by adding local interaction terms to the Lagrangian, one can map the results of one scheme onto those of another. Subsequently, we *renormalize* the theory. This means that all parameters in the Lagrangian are modified by finite corrections, which however may diverge in the limit $\varepsilon \to 0$. If these counter terms have been chosen well, the theory may stay finite and well defined in this limit. In particular, we should have a unitary, causal theory.

Unitarity is only guaranteed if the theory is gauge-invariant. Therefore, one prefers regulator schemes that preserve gauge-invariance throughout. This is what dimensional

---

[8]Only in one case, there is a complication, namely, when there are Adler-Bell-Jackiw anomalies; see Section 8.

regularization often does. In that case, the renormalization procedure respects BRST-invariance, see Subsect. 4.5.

8.   Anomalies

The Sections that follow will (again) be too brief to form a complete text for learning Quantum Field Theory. Our aim is here to give a summary of the features that are all extremely important to understand the general structure of relativistic Quantum Field Theories.

If, for a given theory, no obviously gauge-invariant regularization procedure appears to exist, this might be for a reason: such a theory might not be renormalizable at all. In principle, this could be checked, as follows. One may always decide to use a regularization procedure that does not respect the symmetries one wants, provided that the symmetry can be restored in the limit where the physically observable effects of the regulator go away, such as $\varepsilon \to 0$, or $\Lambda_i \to \infty$, $i > 0$. If a gauge-invariant regulator does exist, but it hasn't yet been explicitly constructed, then we know that it differs from any other regulator by a bunch of finite counter terms. To find such counter terms is not hard, in practice; just add all terms needed to restore BRST invariance of the amplitudes.

But, in case that regulator is not known, how can we then be sure that such terms exist at all? BRST invariance requires the validity of the Slavnov-Taylor identities, but they appear to overdetermine the subtraction terms. This is the way we originally phrased the problem in Ref. [14]. In fact, indeed there may be a clash. If this happens, it is called an *anomaly*.[20]

Actually, the incidence of such anomalies is limited, fortunately. This is because for most theories completely gauge-invariant regulator techniques were found. Dimensional regularization often works. The one case where it does not is when there are *chiral fermions*. Classically, one may separate any fermionic field into a left-handed and a right handed part, as was mentioned in Subsection 4.1:

$$\psi(x) = P_+\psi_L(x) + P_-\psi_R(x) \ ; \qquad P_\pm = \tfrac{1}{2}(1 \pm \gamma^5) \qquad ;$$
$$\gamma_5 = \tfrac{1}{24}\varepsilon_{\mu\nu\alpha\beta}\gamma^\mu\gamma^\nu\gamma^\alpha\gamma^\beta \ . \tag{8.1}$$

Indeed, since $(\gamma^5)^2 = 1$, the operators $P_\pm$ are genuine projection operators: $P_\pm^2 = P_\pm$.

The left- and right sectors of the fermions, see Eq. (4.22), may be separately gauge-invariant, transforming differently under gauge transformations. This, however, requires γ^5 to anti-commute with all other γ^μ, $\mu = 1, \cdots, n$. But, as we see from their definition, Eq. (8.1), γ^5 only anti-commutes with four of the γ^μ, not all n. This is why the contributions from the $-\varepsilon$ remaining dimensions will not be gauge-invariant.

It was discovered by Bell and Jackiw[18], and independently by Adler[19], that no local counter term exists that obeys all symmetry conditions and has the desired dimensionality; Bell and Jackiw tried to use unconventional regulators, but those turned out not to be admissible. The basic culprit is the *triangle diagram*, Fig. 5(a), representing the matrix element of the axial vector current $\overline{\psi}\,\gamma^\mu\gamma^5\,\psi$ in the field of two photons, each being coupled to the vector current $\overline{\psi}\,\gamma^\alpha\,\psi$.
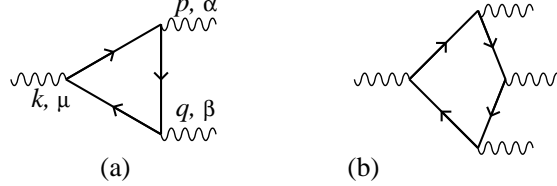

Figure 5: (a) The anomalous triangle diagram. μ, α and β are the polarizations, k, p and $q = k - p$ are the external momenta. (b) An anomalous diagram in non-Abelian theories

For simplicity, we assume here the fermions to be massless. Let us call this amplitude then $\Gamma_\mu^{\alpha,\beta}(p,q)$. It is linearly divergent. Upon regularization, there are two counter terms, or subtraction terms, whose coefficients should be determined, in a correct combination with the finite parts of the amplitude. Limiting ourselves to the correct quantum numbers and dimensions, we find the two quantities,

$$
\begin{aligned}
\delta_1\Gamma_\mu^{\alpha,\beta}(p,q) &= \varepsilon_{\mu\alpha\beta\gamma}p_\gamma \ ; \\
\delta_2\Gamma_\mu^{\alpha,\beta}(p,q) &= \varepsilon_{\mu\alpha\beta\gamma}q_\gamma \ .
\end{aligned}
\tag{8.2}
$$

We can determine their coefficients by applying the condition that the total amplitude be invariant under gauge transformations of the photon field. This implies that the expression must vanish when any of the two photons are longitudinal: $A_\mu = \partial_\mu\Lambda$, which means

$$
p_\alpha\Gamma_\mu^{\alpha,\beta}(p,q) = 0 \ ; \qquad q_\beta\Gamma_\mu^{\alpha,\beta}(p,q) = 0 \ .
\tag{8.3}
$$

Since

$$
\begin{aligned}
p_\alpha\delta_1\Gamma_\mu^{\alpha,\beta}(p,q) &= \quad 0 \ ; & q_\beta\delta_1\Gamma_\mu^{\alpha,\beta}(p,q) &= A_{\mu,\alpha} \ ; \\
p_\alpha\delta_2\Gamma_\mu^{\alpha,\beta}(p,q) &= A_{\mu,\beta} \ ; & q_\beta\delta_2\Gamma_\mu^{\alpha,\beta}(p,q) &= 0 \ ; \\
A_{\mu,\alpha} &= \varepsilon_{\mu\alpha\beta\gamma}p_\gamma q_\beta \ ,
\end{aligned}
\tag{8.4}
$$

this fixes the coefficients in front of $\delta_1\Gamma$ and $\delta_2\Gamma$.

When now we investigate whether this amplitude is also transversal with respect to the axial vector current, we are struck by a surprise. The counter terms, fixed by condition (8.4), also contribute here:

$$k_\mu \delta_1 \Gamma_\mu^{\alpha,\beta}(p,q) = -k_\mu \delta_2 \Gamma_\mu^{\alpha,\beta}(p,q) = A_{\alpha,\beta} \ , \tag{8.5}$$

but they do not cancel against the contribution of the finite part. After imposing gauge-invariance with respect to the two vector insertions, one finds (in the case of a single chiral fermion)

$$k_\mu \Gamma_\mu^{\alpha,\beta}(p,q) = (4\pi^2)^{-1} \varepsilon_{\mu\alpha\beta\gamma} p_\mu q_\gamma \ , \tag{8.6}$$

and this can be rewritten as an effective divergence property of a vector current:

$$\partial_\mu J_\mu^5 = -\frac{iLe^2}{8\pi^2} F_{\mu\nu} \tilde{F}_{\mu\nu} \ , \tag{8.7}$$

where $\tilde{F}_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta} F_{\alpha\beta}$, and it was assumed that the photons couple with charges e.

What is surprising about this is, that the triangle diagram itself, Fig. 5a, appears to be totally symmetric under all permutations, since γ^5 can be permuted to any of the other end-points. Imposing gauge-invariance at two of its three end-points implies breaking of the invariance at the third.

This result is very important. It implies an *induced violation of a conservation law*, apparently to be attributed to the regularization procedure. It also means that it is not possible to couple three *gauge bosons* to such a triangle graph, because this cannot be done in a gauge-invariant way. In most theories, however, we have couplings both to left-handed and to right-handed fermions. Their contributions are of opposite sign, which means that they can cancel out. Therefore, one derives an important constraint on gauge theories with chiral fermions: *The triangle anomalies must cancel out.*

Let the left handed chiral fermions be in representations of the total set of gauge groups that transform as

$$\psi_L^i \to \psi_L^i + i\Lambda^a T^{ai}_{\ j} \psi_L^j \ , \tag{8.8}$$

where Λ is infinitesimal, and T^a are and similarly the ψ_R^i. Define

$$d_L^{abc} = \text{Tr}(T_L^a T_L^b T_L^c + T_L^b T_L^a T_L^c) \ , \tag{8.9}$$

and similarly d_R^{abc}. The anomaly constraint is then

$$\sum d_L^{abc} = \sum d_R^{abc} \ , \tag{8.10}$$

where the sum is over all fermion species in the theory. In the Standard Model, the only contributions could come if either one or all three indices of d^{abc} refer to the $U(1)$

group. One quickly verifies that indeed the $U(1)$ charges of the quarks and leptons are distributed in such a way that (8.10) is completely verified, but *only if the number of quark generations and lepton generations are equal.* In Chapter 10.3, we will see the physical significance of this observation.

Note, that in the non-Abelian case, there are also anomalies in diagrams with 4 external legs, see Fig. 5(b). They arise from the trilinear terms in $F_{\mu\nu}\tilde{F}_{\mu\nu}$ (the quadrilinear terms cancel). These are the *only* cases where the regularization procedure may violate gauge invariance. In diagrams with more loops, or sub diagrams with more external lines, regularization procedures could be found that preserve gauge invariance.

9. Asymptotic freedom

9.1. The Renormalization group

It was observed by Stueckelberg and Peterman[21] in 1953, that, although the perturbative expansion of a theory depends on how one splits up the bare parameters in the Lagrangian into lowest order parameters, and counter terms required for the renormalization, the entire theory should not depend on this. This they interpreted as an invariance, and the action of replacing parameters from lowest order to higher order corrections as a group operation. One obtains the 'Renormalization Group'.

There is only one instance where such transformtions really matter, and that is when one compares a theory at one mass- or distance-scale to the same theory at a different scale. A scale transformation must be associated with a replacement of counter terms. Thus, physicists began to identify the notion of a scale transformation as a 'renormalization group transformation'.

Gell-Mann and Low[22] observed that this procedure can be used to derive the small-distance behavior of QED. One finds that the effective fine-structure constant depends on the scale μ, described by the equation

$$\frac{\mu^2 \mathrm{d}}{\mathrm{d}\mu^2}\alpha(\mu) = \beta(\alpha) = \frac{\alpha^2 N}{3\pi} + \mathcal{O}(\alpha^3) \ , \tag{9.1}$$

where N is the number of charged fermion types. As long as $\alpha(\mu)$ stays small, so that the $\mathcal{O}(\alpha^3)$ terms can be neglected, we see that its μ-dependence is

$$\alpha(\mu) = \frac{\alpha_0}{1 - (\alpha_0 N/3\pi)\log(\mu^2/\mu_0^2)} \ , \qquad \text{if} \quad \alpha(\mu_0) = \alpha_0 \ . \tag{9.2}$$

Things run out of control when μ reaches values comparable to $\exp(3\pi/2N\alpha_0)$, but, at least in the case of QED, where $\alpha_0 \approx 1/137$, this mass scale is so large that in practice no

problems are expected. The pole in Eq. (9.2) is called the Landau pole; Landau concluded that quantum field theories such as QED have no true continuum limit because of this pole. Gell-Mann and Low suspected, however, that $\beta(\alpha)$ might have a zero at some large value of α, so that, at high values of μ, α approaches this value, but does not exceed this stationary point.

What exactly happens at or near the Landau pole, cannot be established using perturbation expansion alone, since this will depend on *all* higher order terms in Eq. (9.1); in fact, it is not even known whether Quantum Field Theory can be reformulated accurately enough to decide. The question, however, might be not as important as it seems, since the Landau pole will be way beyond the Planck mass, where we know that gravitational terms will take over; it will be more important to solve Quantum Gravity first.

An entirely different situation emerges in theories where the function $\beta(\lambda)$ is negative. It was long thought that this situation can never arise, unless the coupling strength λ itself is given the wrong sign (the sign that would render the energy density of the classical theory unbounded from below), but this turns out only to be the case in theories that only contain scalar and spinor fields. If there is a *non-Abelian* Yang-Mills component in the theory, negative β functions do occur. In the simplest case, an $SU(2)$ gauge theory with N_f fermions in the elementary doublet representation, the beta function is

$$\frac{\mu^2 \mathrm{d}}{\mathrm{d}\mu^2} g^2(\mu) = \beta(g^2) = \frac{N_f - 11}{24\pi^2} g^4(\mu) + \mathcal{O}(g^6) \ , \tag{9.3}$$

so, as long as $N_f < 11$ we have that the coupling strength $g(\mu)$ tends to *zero*, logarithmically, as $\mu \to \infty$. This feature is called *asymptotic freedom*. In an $SU(N_c)$ gauge theory, the β function is proportional to $N_f - \frac{11}{2} N_c$, so, with the present number of $N_f = 6$ quark flavors, QCD ($N_z = 3$) is asymptotically free. In line with a notation often used, the subscript c here stands for "colour"; in QCD, the number of colours is $N_c = 3$.

9.2. An algebra for the beta functions

In theories with gauge fields, fermions, and scalars, the situation is more complex. A general algorithm for the beta functions has been worked out. The most compact notation for the general result can be given by writing how the entire Lagrange density $\mathcal{L}$ scales under a scale transformation. Let the Lagrangian $\mathcal{L}$ be

$$\begin{aligned} \mathcal{L} &= -\tfrac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu} - \tfrac{1}{2}(D_\mu \phi_i)^2 - V(\phi) - \overline{\psi}_i \gamma_\mu D_\mu \psi_i \\ &\quad - \overline{\psi}_i \Big(S_{ij}(\phi) + i\gamma_5 P_{ij}(\phi) \Big) \psi_j \ , \end{aligned} \tag{9.4}$$

where the covariant derivatives are defined as follows:[9]

$$D_\mu \phi_i \equiv \partial_\mu \phi_i + i T^a_{ij} A^a_\mu \phi_j \ ; \qquad D_\mu \psi_i = \partial_\mu \psi_i + i U^a_{ij} A^a_\mu \psi_j \ , \tag{9.5}$$

---

[9] T and U are hermitean, but since ϕ is real, the elements of T must be imaginary.

and the structure constants f^{abc} are defined by

$$[T^a, T^b] = -if^{abc}T_c , \qquad (9.6)$$

so that

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu . \qquad (9.7)$$

We split the fermions into right- and left-handed representations, so that

$$U^a = U^a_L P_L + U^a_R P_R ; \qquad P_L = \frac{1+\gamma_5}{2} , \quad P_R = \frac{1-\gamma_5}{2} . \qquad (9.8)$$

The functions $S(\phi)$ and $P(\phi)$ are at most linear in ϕ and $V(\phi)$ is at most quartic. The Lagrangian (9.4) is the gauge-invariant part; we do not write the gauge-fixing part or the ghost; the final result will not depend on those details.

The result of an algebraical calculation is that

$$16\pi^2 \frac{\mu^2 \mathrm{d}\mathcal{L}}{\mathrm{d}\mu^2} =$$

$$G^a_{\mu\nu} G^b_{\mu\nu} [-\frac{11}{12}C_1^{ab} + \frac{1}{24}C_2^{ab} + \frac{1}{6}C_3^{ab}] - \Delta V - \overline{\psi}(\Delta S + i\gamma_5 \Delta P)\psi , \qquad (9.9)$$

in which

$$C_1^{ab} = f^{apq}f^{bpq} , \qquad (9.10)$$
$$C_2^{ab} = \mathrm{Tr}\,(T^a T^b) , \qquad (9.11)$$
$$C_3^{ab} = \mathrm{Tr}\,(U^a_L U^b_L + U^a_R U^b_R) , \qquad (9.12)$$
$$\Delta V = \tfrac{1}{4}V_{ij}^2 - \tfrac{3}{2}V_i(T^2\phi)_i + \tfrac{3}{4}(\phi T^a T^b \phi)^2$$
$$+ \phi_i V_j \mathrm{Tr}\,(S_{,i}S_{,j} + P_{,i}P_{,j}) - \mathrm{Tr}\,(S^2 + P^2)^2 + \mathrm{Tr}\,[S, P]^2 , \qquad (9.13)$$

where

$$V_i \equiv \frac{\partial V(\phi)}{\partial \phi_i} ; \quad S_{,i} \equiv \frac{\partial S}{\partial \phi_i} , \qquad \text{etc.,} \qquad (9.14)$$

and writing $S + iP = W$, one finds ΔS and ΔP from

$$\Delta W = \tfrac{1}{4}W_i W_i^* W + \tfrac{1}{4}W W_i^* W_i + W_i W^* W_i$$
$$- \tfrac{3}{2}(U_R^2 W) - \tfrac{3}{2}W(U_L)^2 + W_i \phi_j \mathrm{Tr}\,(S_i S_j + P_i P_j) . \qquad (9.15)$$

This expression does not include information on how fields ϕ_i and ψ_i transform under scaling. The fields are not directly observable.

This algebraic expression can be used to find how, in general, coupling strengths run under rescalings of the momenta. It is an interesting exercise to work out what the conditions are for asymptotic freedom, that is, for all coupling strengths to run to zero at infinite momentum. In general, one finds that scalar fields can only exist if there are also gauge fields and fermions present; the latter must be in sufficiently high representations of the gauge group.

58

10.   Topological Twists

The Lagrangian (9.4) is the most general one allowed if we wish to limit ourselves to coupling strengths that run *logarithmically* under rescalings of the momenta, see for instance Eq. (9.2). Such theories have a domain of validity that range over exponentially large values of the momenta (in principle over all momenta if the theory is asymptotically free). The most striking feature of this general Lagrangian is that it is topologically highly non-trivial. Locally stable field configurations may exist that have some topological twist in them. In particular, this can be made explicit in the case of a Brout-Englert-Higgs mechanism. Here, these twists can already be seen at the classical level (i.e., ignoring quantum effects).

If we say that a scalar field ϕ_i has a vacuum expectation value, then this means that we perform our perturbation expansion starting with a field value of the form $\phi_i = (F, 0, \cdots)$ in the vacuum, after which field fluctuations $\delta\phi$ around this value are assumed to be small. One assumes that the potential $V(\phi)$ has its minimum there. This may appear to violate gauge-invariance, if ϕ_i transform into each other under local gauge transformations, but strictly speaking the phrase "spontaneous breakdown of local gauge symmetry" is inappropriate, because it may also simply mean that we choose a gauge condition. It is however a fact that the spectrum of physical particles comes out to be altogether different if we perturb around $\phi_i = 0$, so this 'Higgs mode' is an important notion in any case.

10.1.   Vortices

If the Higgs field has only two real components (such as when $U(1)$ is broken into the identity group), one may consider a configuration where this field makes a full twist over 360° when following a closed curve. Inside the curve there must be a zero. The zeros must form a curve themselves, and they cost energy. This is the Abrikosov vortex. Away from its center, one may transform ϕ_i back to a constant value, but this generates a vector potential $A_\mu(x)$, obeying

$$\oint A_\mu \mathrm{d}x^\mu = \frac{2\pi}{e} \ , \tag{10.1}$$

which means that this vortex carries an amount of magnetic flux, of magnitude exactly $2\pi/e$. Apparently, in this model, magnetic field lines condense into locally stable vortices.[23] This is also what happens to magnetic fields inside a superconductor.

10.2.   Magnetic Monopoles

Something similar may happen if the Higgs field has three real components. In that case, one can map the S_2 sphere of minima of $V(\phi)$, onto a sphere in 3-space. There will be isolated zeros inside this sphere. These objects behave as locally stable particles. If one tries to transform the field locally to a constant value, one finds that a vector potential again may emerge.

If, for example, in an $SU(2)$ theory, a Higgs in the adjoint representation (which has 3 real components) breaks the gauge group down to $U(1)$, then one finds the vector potential of an isolated magnetic source inside the sphere. This means that the source is a magnetic monopole with magnetic charge $g_m = \frac{4\pi}{e}$, where e is the original coupling strength of the $SU(2)$ theory. Indeed, Dirac[24] has derived, back in 1931, that magnetic charges g_m and electric charges q must obey the *Dirac quantization condition*

$$q\, g_m = 2\pi n \; . \tag{10.2}$$

Apparently, for the monopole in this model, $n = 2$. However, it is easy to introduce particles in the elementary representation, which have $q = \frac{1}{2}e$; these then saturate the Dirac condition (10.2).

Dirac could not say much about the mass of his magnetic monopoles. In the present theories, however, the mass is calculable. In general, the magnetic monopole mass turns out to be the mass of an ordinary particle divided by a number of the order of the gauge coupling strength squared.

Careful analysis of the existing Lie groups and the way they may be broken spontaneously into one or more subgroups $U(1)$, reveals a general feature: *Only if the underlying gauge group is compact, and has a compact covering group, must electric charges in the $U(1)$ gauge groups be quantized* (otherwise, it would not be forbidden to add arbitrary real numbers to the $U(1)$ charges), *and whenever the covering group of the underlying gauge group is compact, magnetic monopole solutions can be constructed.* Apparently, whenever the gauge group structure provides for a compelling reason for electric charges to be quantized, the existence of magnetic monopole solutions is guaranteed. Thus, assuming that Nature has compelling reasons for the charge units of electrons and protons to be equal, and quantized into multiples of e, we must assume that magnetic monopole solutions must exist. However, in most 'Grand Unification Schemes', the relevant mass scale is many orders of magnitude higher than the mass scale of particles studied today, so the monopoles, whose mass is that divided by a coupling strength squared, are even heavier.

From the structure of the Higgs field of a monopole, one derives that the system is invariant under rotations provided that rotations are associated with gauge rotations. A consequence of this is, that elementary particles with half-odd isospin, when bound to

a monopole, produce bound states with half-odd integer *orbital* angular momentum[25]. What is strange about this, is that such particles should develop Dirac statistics. Indeed, one can derive that both the spin and the statistics of bound states of electric and magnetic charges, flip from Bose-Einstein to Fermi-Dirac or back[26] if they form odd values of the Dirac quantum n (Eq. 10.2).

10.3. Instantons

A Higgs field with two real components gives rise to vortices, a Higgs with three components gives magnetic monopoles, so what do we get if a Higgs field has four real components? This is the case if, for instance, $SU(2)$ is broken spontaneously into the identity by a Higgs in the fundamental representation (two complex $=$ 4 real components). The topologically stable objects one finds are stable points in four-dimensional space-time. They represent *events*, and, referring to their particle-like appearance, the resulting solutions (in Euclidean space) were called 'instantons'. Because this Higgs field, in the case of $SU(2)$, breaks the gauge symmetry completely, one can argue that this solution is also topologically stable in pure gauge theories, without a Higgs mechanism at all. Far from the origin, the vector potential field is described as a local gauge rotation of the value $A_\mu^a(x) = 0$. The gauge rotation in question, $\Omega(x)$, is described by noting that the $SU(2)$ matrices form an S_3 space, *i.e.*, the three dimensional surface of a sphere in four dimensions. Mapping this S_3 one-to-one onto the boundary of some large region in (Euclidean) space-time, gives the field configuration of an instanton.

It was noted by Belavin, Polyakov, Schwarz and Tyupkin[27] (who also were the first to write down this solution) that this solution has a non-vanishing value of

$$\int \mathrm{d}^4x \, F_{\mu\nu}^a \tilde{F}_{\mu\nu}^a = \frac{32\pi^2}{g^2} \ . \tag{10.3}$$

The integrand is the divergence of a current:

$$F_{\mu\nu}^a \tilde{F}_{\mu\nu}^a = \partial_\mu K_\mu \ ; \qquad K_\mu = 2\varepsilon_{\mu\nu\alpha\beta}A_\nu^a(\partial_\alpha A_\beta^a + \tfrac{1}{3}gf^{abc}A_\alpha^b A_\beta^c) \ , \tag{10.4}$$

the so-called Chern-Simons current. This current, however, is not gauge-invariant, which is why it does not vanish at infinity. It does vanish after the gauge transformation $\Omega(x)$ that replaces A_μ^a at infinity by 0. Eq. (10.3) is most easily derived by using this Chern-Simons current. It so happens that the instanton is also a solution of the equation

$$F_{\mu\nu} = \tilde{F}_{\mu\nu} \ , \tag{10.5}$$

so that we also find the action to be given by $-8\pi^2/g^2$.

In a pure gauge theory (one without fermions), instantons can be interpreted as representing *tunneling transitions*. In ordinary Quantum Mechanics, tunneling is an exponentially suppressed transition. The exponential suppression is turned into an oscillating

expression if we replace time t by an imaginary quantity $i\tau$. The oscillating exponent is the action of a classical transition in imaginary time. One may also say that a tunneling transition can be described by a classical mechanical transition if the potential $V(\vec{q})$ is replaced by $2E - V(\vec{q})$, where E is the energy. The classical action then represents the quantity in the exponent of the (exponentially suppressed) tunneling transition.

The above substitution is exactly what one gets by replacing time t by $i\tau$. In relativistic Quantum Field Theory, this is also exactly the Wick rotation from Minkowski space-time into Euclidean space-time. In short, instantons represent tunneling that is associated with the suppression factor $e^{-8\pi^2/g^2}$.

The transition can be further understood by formulating a gauge theory in the temporal gauge, $A_0 = 0$. In this gauge, there is a residual invariance under gauge transformations $\Lambda(\mathbf{x})$ that are time-independent. All 'physical states', therefore, come as representations of this local gauge group. Normally, however, we restrict ourselves to the trivial representation, $\Omega|\psi\rangle = |\psi\rangle$, where $\Omega = e^{i\int \Lambda(\mathbf{x})\mathrm{d}^3\mathbf{x}}$, because this configuration is conserved in time, and because any other choice would violate Lorentz invariance. However, closer analysis shows that one only has to impose this constraint for those gauge transformations that can be continuously reached from the identity transformation. This is *not* the case for transformations obtained by mapping the S_3 space of the $SU(2)$ transformations non-trivially onto three-space $\mathbb{R}_3$. These transformations form a discrete set, characterized by the integers $k = 0, \pm 1, \pm 2, \ldots$. Writing

$$\Omega_k(\mathbf{x}) = \Omega_1(\mathbf{x})^k , \qquad \Omega_k|\psi\rangle = e^{i\theta k}|\psi\rangle , \tag{10.6}$$

we find that the tunneling transitions described by instantons cause an exponentially suppressed θ dependence of physical phenomena in the theory. Since, under parity transformations P, the angle θ turns into $-\theta$, a non-vanishing θ also implies an explicit parity (eventually, PC) violation of the strong interactions.

In the presence of fermions, the situation is altogether different. Due to the chiral anomaly, we have for the current of chiral fermions $J_\mu^5(x)$, the equation (8.7). The total number of chiral fermions, $Q^5 = \int \mathrm{d}^3\mathbf{x} J_0^5(\mathbf{x})$ changes by one unit due to an instanton: $\Delta Q^5 = \pm 1$. This can be understood by noting that the Dirac equation for massless, chiral fermions has one localized solution in the Euclidean space of an instanton. In Minkowski space-time, this solution turns into a state that describes a chiral fermion either disappearing into the Dirac sea, or emerging from it, so that, indeed, the number of particles minus anti-particles changes by one unit for every chiral fermion species. If left- and right handed fermions are coupled the same way to the gauge field, as in QCD, the instanton removes a left-handed fermion and creates a right-handed one, or, in other words, it flips the chirality. This $\Delta Q^5 = \pm 2$ event has exactly the quantum numbers of a mass term for the Goldstone boson that would be associated to the conservation of chiral charge, the η particle. This explains why the η particle is considerably heavier than the pions, which have lost most of their mass due to chiral symmetry of the quarks.[28]

What one concludes from the study of instantons is that QCD, the theory for the strong interactions, neatly explains the observed symmetry structure of the hadron spectrum, including the violation of chiral charge conservation that accounts for the η mass.

In the electro-weak sector, one also has instantons. We now see that the cancellation of the anomalies in the quark and the lepton sector implies an important property of the electro-weak theory: since the anomalies do not respect gauge-invariance of the quark sector alone, quarks can be shown not to be exactly conserved. One finds that instantons induce baryon number violating events: three baryons (nine different quarks all together) may transmute into three anti-leptons, or vice versa.

11.   Confinement

An important element in the Standard Model is the gauge theory for the strong interactions, based on the gauge group $SU(3)$. Quarks are fermions in the elementary representation of $SU(3)$. The observed hadronic particles all are bound states of quarks and/or anti-quarks, in combinations that are gauge-invariant under $SU(3)$. An important question is: what is the nature of the forces that binds these quarks together? We have seen that vortex solutions can be written down that would cause an interesting force pattern among magnetic monopoles: in a Higgs theory with magnetic monopoles, these monopoles could be bound together with Abrikosov vortices.

Indeed, this would be a confining force: every magnetic monopole must be the end point of a vortex, whose other end point is a monopole of opposite magnetic charge. Indeed, the confinement would be *absolute*: isolated monopoles cannot exist. It was once thought that, therefore, quarks must be magnetic monopoles. This, however, would be incompatible with the finding that quarks only interact weakly at small distances, magnetic charges being always quite strong. A more elegant idea is that the binding force forms electric rather than magnetic vortices. An electric vortex can be understood as the *dual transformation* of a magnetic vortex. It comes about when the Brout-Englert-Higgs mechanism affects freely moving magnetically charged particles. Further analytic arguments, as well as numerical investigations, have revealed that indeed such objects are present in QCD, and that the Higgs mechanism may occur in this sector. Let us briefly explain the situation in words.

11.1.   The maximally Abelian gauge

A feature that distinguishes non-Abelian gauge theories from Abelian ones, is that a reference frame for the gauge choice, the gauge condition, can partly be fixed locally in terms of the pure gauge fields alone; noticing that the covariant field strengths $G_{\mu\nu}$ transform as the adjoint representation, one may choose the gauge such that one of these components,

63

say G_{12}, is diagonal. This then removes the non-Abelian part of the gauge group, but the diagonal part, called the Cartan subgroup, remains. In this way, a non-Abelian gauge theory turns into an Abelian one. A slightly smarter, but non-local gauge that does the same is the condition that $\sum_{i \neq j}(A_{\mu j}^{\ i})^2$ is minimized. It is called the maximally Abelian gauge.

However, such a gauge choice does produce singularities. These typically occur when two eigenvalues of G_{12} coincide. It is not difficult to convince oneself that these singularities behave as particles, and that these particles carry magnetic charges with respect to the Cartan subgroup. *Absolute confinement occurs as soon as these magnetically charged particles undergo a Brout-Englert-Higgs mechanism.*

Although this still is the preferred picture explaining the absolute nature of the quark confining force, it may be noted that the magnetically charged particles do not have to be directly involved with the confinement *mechanism*. Rather, they are indicators. This, we deduce from the fact that confinement also occurs in theories with a very large number N_c of colors; in the limit $N_c \to \infty$, magnetically charged particles appear to be suppressed in the perturbative regime, but the electric vortices are nevertheless stable. The strength of a vortex is determined by its finite width, and this width is controlled by the lightest gluonic state, the 'glueball'. At distance scales large compared to the inverse mass of the lightest glueball, an electric vortex cannot break.

Confinement is a condensation phase that is a logical alternative of the Brout-Englert-Higgs phase. In some cases, however, these two phases may coexist. An example of such a coexistence is the $SU(2)$ sector of the Standard Model. Conventionally, this sector is viewed as a prototype of the Higgs mechanism, but it so happens that the $SU(2)$ sector of the Standard Model can be treated exactly like the colour $SU(3)$ sector: as if there is confinement. To see this, one must observe that the Higgs doublet field can be used to fix the $SU(2)$ sector of the gauge group unambiguously. This means that all physical particles can be connected to gauge-invariant sources by viewing them as gauge-invariant bound states of the Higgs particle with the other elementary doublets of the model. For instance, writing the Higgs doublet as $\phi_a = \left(\begin{smallmatrix} F \\ 0 \end{smallmatrix}\right) + \tilde{\phi}_a$, and the lepton doublet as ψ_a, the electron is seen to be associated to the 'baryonic' field $\varepsilon^{ab}\phi_a\psi_b$, the neutrino is $\phi^{*a}\psi_a$, the Z_0 boson is $\phi^{*a}D_\mu\phi_a$, and so on.

Theories in which the confinement phase is truly distinct from the Higgs phase are those where the Higgs field is not a one-to-one representation of the gauge group, such as the adjoint representation of $SU(2)$.

12. Outlook

Quantum Field Theory has reached a respectable status as an accurate and well-studied description of sub-atomic particles. From a purely mathematical point of view, there are some inherent limitations to the accuracy by which it defines the desired amplitudes, but in nearly all conceivable circumstances, its intrinsic accuracy is much higher than what can be reached in experiments. This does not mean that we can reach such accuracy in real calculations, which more often than not suffer from technical limitations, particularly where the interactions are strong, as in QCD. In this domain, there is still a need for considerable technical advances.

12.1. Naturalness

When the Standard Model, as known today, is extrapolated to energy domains beyond approximately 1 TeV, a difficulty is encountered that is not of a mathematical nature, but rather a physical one: it becomes difficult to believe that it represents the real world. The bare Lagrangian, when considered on a very fine lattice, is required to have parameters that must be tuned very precisely in order to produce particles such as the Higgs particle and the weak vector bosons, whose masses are much *less* than 1 TeV. This fine-tuning is considered to be *unnatural*. In a respectable physical theory, such a coincidence is not expected. With some certainty, one can state that the fundamental laws of Nature must allow for a more elegant description at high energies than a lattice with such fine-tuning. What is generally expected is either a new symmetry principle or possibly a new regime with an altogether different set of physical fields.

A candidate for a radically different regime is the so-called *technicolour* theory, a repetition of QCD but with a typical energy scale of a TeV rather than a GeV. The quarks, leptons and Higgs particles of the Standard Model would then all turn out to be the hadrons of this technicolour theory. Different gauge groups could replace $SU(3)$ here. However, according to this scheme, a new strong interaction regime would be reached, where perturbation expansions used in the weak sector of the Standard Model would have to break down. As precision measurements and calculations continue to confirm the reliability of these perturbation expansions, the technicolour scenario is considered to be unlikely.

12.2. Supersymmetry

A preferred scenario is a simple but beautiful enhancement of the symmetries of the Standard Model: supersymmetry. This symmetry, which puts fermions and bosons into single multiplets, does not really modify the fundamental aspects of the theory. But

it does bring about considerable simplifications in the expressions for the amplitudes, not only in the perturbative sector, but also, in many cases, it allows us to look deeper into the non-perturbative domains of the theories. There is a vast amount of literature on supersymmetry, but some aspects of it are still somewhat obscure. We would like to know more about the physical origin and meaning of supersymmetry, as well as the mechanism(s) causing it to be broken — and made almost invisible — at the domain of the Standard Model that is today accessible to experimental observation.

12.3.   Resummation of the Perturbation Expansion

The perturbation expansion in Quantum Field Theory is almost certain to be divergent for any value of the coupling parameter(s). A simple argument for its divergence has been put forward by Dyson[29]: imagine that in the theory of QED there were a bound ε such that, whenever $|\alpha| < \varepsilon$, where α is the fine-structure constant, perturbation expansions would converge. Then it would converge for some negative real value of α. However, one can easily ascertain that for any negative value of α, the vacuum would be unstable: vacuum fluctuations would allow large numbers of electrons to be pair-created, and since like charges attract, highly charged clouds of electrons could have negative energies.

Theories with asymptotic freedom may allow for a natural way to re-sum the perturbation series, by first solving the theory at high energy with extreme precision, after which one has to integrate the Schrödinger equation to obtain the physical amplitudes at lower energy. Such a program has *not* yet been carried out, because integrating these Schrödinger equations is beyond our present capabilities, but one may suspect that, as a matter of principle, it should be possible. Theories that are not asymptotically free may perhaps allow for more precise treatments if an ultra-violet fixed point can be established.

The extent of the divergence of the perturbation expansion can be studied or predicted. This one does using the Borel resummation technique. An amplitude

$$\Gamma(\lambda) = \sum_{n=1}^{\infty} a_n \lambda^n \ , \tag{12.1}$$

can be rewritten as

$$\begin{aligned} \Gamma(\lambda) &= \int_0^{\infty} B(z) e^{-z/\lambda} \mathrm{d}z \ , \\ B(z) &= \sum_{k=0}^{\infty} a_{k+1} z^k / k! \ . \end{aligned} \tag{12.2}$$

The series for $B(z)$ is generally expected to have a finite radius of convergence. If $B(z)$ can be analytically extended to the domain $0 \leq z < \infty$, then that (re-)defines our amplitude. In general, however, one can derive that $B(z)$ must have singularities on the

66

real axis, for instance where z corresponds to the action of instantons or instanton pairs. In addition, singularities associated to the infrared and/or ultraviolet divergences of the theory are expected. Sometimes, these different singularities interfere.

12.4.   General Relativity and Superstring Theory

It is dubious, however, whether the issue of convergence or divergence of the perturbation expansion is of physical relevance. We *know* that Quantum Field Theory cannot contain the entire truth concerning the sub-atomic world; the gravitational force is guaranteed not to be renormalizable, so at those scales where this force becomes comparable to the other forces, the so-called Planck scale, a radically new theory is called for. *Superstring Theory* is presently holding the best promise to evolve into such a theory. With this theory, physicists are opening a new chapter, where we leave conventional Quantum Field Theory, as described in this paper, behind. In its present form, superstring theory appears to have turned into a collection of wild ideas called M-theory, whose foundations are still extremely shaky. Some of the best minds of the world are competing to turn this theory into something that can be used to provide for reliable predictions and that can be taught in a text book, but this has not yet been achieved.

References

[1] A. Pais. *Inward bound: of matter and forces in the physical world,* Oxford University Press 1986. See also: R.P.Crease and C.C. Mann, *The Second Creation: makers of the revolution in twentieth-century physics,* New York, Macmillan, 1986.

[2] B. de Wit and J. Smith, *Field Theory in Particle Physics*, North Holland, 1986, ISBN 0444 86 999 9; I.J.R. Aitchison and A.J.G. Hey, *Gauge Theories in Particle Physics*, Adam Hilger, 1989, ISBN 0-85274-328-9; L.H. Ryder, *Quantum Field Theory*, Cambridge Univ. Press, 1985, 1996, ISBN 0 521 47814 6; C. Itzykson and J.-B. Zuber, *Quantum Field Theory*, McGraw Hill, New York, 1980; T.-P. Cheng and L.-F. Li, *Gauge theory of elementary particle physics*, Clarendon Press, Oxford, 1984, ISBN 0 19 851961 3 (Pbk).

[3] C.N. Yang and R.L. Mills, *Physical Review* **95** (1954) 631.

[4] B.S. DeWitt, *Physical Review Letters* **12** (1964) 742.

[5] B.S. DeWitt, *Physical Review* **160** (1967) 1113; ibid. **162** (1967) 1195, 1239.

[6] L.D. Faddeev and V.N. Popov, *Physics Letters* **25B** (1967) 29; L.D. Faddeev, Theoretical and Mathematical Physics **1** (1969) 3 (in Russian) L.D. Faddeev, Theoretical

and Mathematical Physics **1** (1969) 1 (Engl. transl) L. Faddeev, L. Takhtajan, Soviet Journal of Mathematics 24 (1984) 241.

[7] C. Becchi, A. Rouet and R. Stora, *Communications in Mathematical Physics* **42** (1975) 127; id., *Annals of Physics* (N.Y.) **98** (1976) 287.

[8] I.V. Tyutin, Lebedev Prepr. FIAN 39 (1975), unpubl.

[9] G. 't Hooft and M. Veltman, *Nuclear Physics* **B44** (1972) 189.

[10] F. Englert and R. Brout, *Physical Review Letters* **13** (1964) 321.

[11] P.W. Higgs, *Physics Letters* **12** (1964) 132; *Physical Review Letters* **13** (1964) 508; *Physical Review* **145** (1966) 1156.

[12] L. Hoddeson *et al*, *The Rise of the Standard Model*, Particle Physics in the 1960s and 1970s, Cambridge Univ. Press, 1997, ISBN 0-521-57816-7.

[13] G. 't Hooft and M. Veltman, *DIAGRAMMAR*, CERN Report 73/9 (1973), reprinted in *Particle Interactions at Very High Energies*, Nato Adv. Study Inst. Series, Sect. B, vol. **4b**, p. 177; reprinted in G. 't Hooft, *Under the Spell of the Gauge Principle*, Adv Series in Math. Physics Vol **19**, World Scientific, Singapore, 1994, p. 28.

[14] G. 't Hooft, *Nuclear Physics* **B33** (1971) 173.

[15] G. 't Hooft and M. Veltman, *Nuclear Physics* **B50** (1972) 318.

[16] A. Slavnov, *Theoretical and Mathematical Physics* **10** (1972) 153 (in Russian), *Theoretical and Mathematical Physics* **10** (1972) 99 (Engl. Transl.)

[17] J.C. Taylor, *Nuclear Physics* **B33** (1971) 436.

[18] J.S. Bell and R. Jackiw, *Il Nuovo Cimento* **60A** (1969) 47.

[19] S.L. Adler, *Physical Review* **177** (1969) 2426; S.L. Adler and W.A. Bardeen, *Physical Review* **182** (1969) 1517; W.A. Bardeen, *Physical Review* **184** (1969) 1848.

[20] R. Jackiw, *Diverse topics in Theoretical and Mathematical Physics*, World Scientific 1995, Chapter 1.

[21] E,C,G. Stueckelberg and A. Peterman, *Helvetica Physica Acta* **26** (1953) 499; N.N. Bogoliubov and D.V. Shirkov, *Introduction to the theory of quantized fields* (Interscience, New Yark, 1959).

[22] M. Gell-Mann and F. Low, *Physical Review* **95** (1954) 1300.

[23] H.B. Nielsen and P. Olesen, *Nuclear Physics* **B61** (1973) 45.

[24] P.A.M. Dirac, *Proceedings of the Royal Society* **A133** (1931) 60; *Physical Review* **74** (1948) 817. 9.

[25] P. Hasenfratz and G. 't Hooft , *Physical Review Letters* **36** (1976) 1119.

[26] A.S. Goldhaber, *Physical Review Letters* **36** (1976) 1122.

[27] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Y.S. Tyupkin, *Physics Letters* **59 B** (1975) 85.

[28] G. 't Hooft, *Physics Reports* **142** (1986) 357.

[29] F. Dyson, *Physical Review* **85** (1952) 631.