Published by CERN in the Proceedings of the 2015 CERN–Latin-American School of High-Energy Physics, Ibarra, Ecuador, 4 – 17 March 2015, edited by M. Mulders and G. Zanderighi, CERN-2016-005 (CERN, Geneva, 2016)

Introduction to the Standard Model of the Electro-Weak Interactions

J. Iliopoulos
Laboratoire de Physique Théorique de L'Ecole Normale Supérieure, Paris, France

Abstract
These lectures notes cover the basic ideas of gauge symmetries and the phenomenon of spontaneous symmetry breaking which are used in the construction of the Standard Model of the Electro-Weak Interactions.

Keywords
Lectures; Standard Model; electroweak interaction; gauge theory; spontaneous symmetry breaking; field theory.

1 Introduction

These are the notes from a set of four lectures that I gave at the 2015 European Organization for Nuclear Research (CERN)–Latin-American School of High-Energy Physics as an introduction to more specialized lectures. With minor corrections, they follow the notes of the lectures I gave at the 2012 CERN Summer School. In both cases, the students were mainly young graduate students doing experimental high-energy physics. They were supposed to be familiar with the phenomenology of particle physics and to have a working knowledge of quantum field theory and the techniques of Feynman diagrams. The lectures were concentrated on the physical ideas underlying the concept of gauge invariance, the mechanism of spontaneous symmetry breaking, and the construction of the Standard Model. Although the methods of computing higher-order corrections and the theory of renormalization were not discussed at all in the lectures, the general concept of renormalizable versus non-renormalizable theories was supposed to be known. Nevertheless, for the benefit of the younger students, a special lecture on the physical principles of renormalization theory was included. It is given as an appendix in these notes. The plan of the notes follows that of the lectures with five sections:

– a brief summary of the phenomenology of the electromagnetic and the weak interactions;
– gauge theories, Abelian and non-Abelian;
– spontaneous symmetry breaking;
– the step-by-step construction of the Standard Model;
– the Standard Model and experiment.

It is generally accepted that progress in physics occurs when an unexpected experimental result contradicts the established theoretical beliefs. As Feynman put it "progress in physics is to prove yourself wrong as soon as possible". This has been the rule in the past, but there are exceptions. The construction of the Standard Model is one of them. In the late 1960s, weak interactions were well described by the Fermi current $\times$ current theory and there was no compelling experimental reason to want to change it: the problems were theoretical. It was only a phenomenological model which, in technical language, was non-renormalizable. In practice, this meant that any attempt to compute higher-order corrections in the standard perturbation theory would give meaningless, divergent results. So the motivation for changing the theory was for aesthetic rather than experimental reasons: it was the search for mathematical consistency and theoretical elegance. In fact, at the beginning, the data did not seem to support the theoretical speculations. Although the history of these ideas is a fascinating subject, I decided not to follow the historical evolution which would have taken more than four lectures to develop. I start instead from the experimental data known at present and show that they point unmistakably to what is known as the Standard Model. In the last section, I recall its many experimental successes.

https://doi.org/10.5170/CERN-2016-005.1

Table 1: Our present ideas on the structure of matter. Quarks and gluons do not exist as free particles and the graviton has not yet been observed.

Table of elementary particles		
	Quanta of radiation	
Strong interactions		Eight gluons
Electromagnetic interactions		Photon (γ)
Weak interactions		Bosons W^+ , W^- , Z^0
Gravitational interactions		Graviton (?)
	Matter particles	
	Leptons	Quarks
1st family	ν_e , e^-	u_a , d_a , $a = 1, 2, 3$
2nd family	ν_μ , μ^-	c_a , s_a , $a = 1, 2, 3$
3rd family	ν_τ , τ^-	t_a , b_a , $a = 1, 2, 3$
	Higgs boson	

2 Phenomenology of the electro-weak interactions: a reminder

2.1 The elementary particles

The notion of an 'elementary particle' is not well defined in high-energy physics. It evolves with time following progress in experimental techniques which, by constantly increasing the resolution power of our observations, have shown that systems that were believed to be 'elementary' are in fact composed of smaller constituents. So, in the last century we went through the chain:

molecules $\rightarrow$ atoms $\rightarrow$ electrons + nuclei $\rightarrow$ electrons + protons + neutrons $\rightarrow$ electrons + quarks $\rightarrow$???

There is no reason to believe that there is an end to this series and, even less, that this end has already been reached. Table 1 summarizes our present knowledge, and the following remarks can be made.

– All interactions are produced by the exchange of virtual quanta. For the strong, electromagnetic, and weak interactions they are vector (spin-one) fields, whereas the graviton is assumed to be a tensor, spin-two field. We shall see in these lectures that this property is well understood in the framework of gauge theories.
– The constituents of matter appear to all be spin one-half particles. They are divided into quarks, which are hadrons, and 'leptons' which have no strong interactions. No deep explanation is known either for their number (why three families?) or for their properties, such as their quantum numbers. We shall come back to this point when we discuss the gauge-theory models. In the framework of some theories that go beyond the Standard Model, such as supersymmetric theories, we can find particles of zero spin among the matter constituents.
– Each quark species, called 'flavour', appears in three forms, often called 'colours' (no relation to the ordinary sense of either word).
– Quarks and gluons do not appear as free particles. They form a large number of bound states, known as the hadrons. This property of 'confinement' is one of the deep unsolved problems in particle physics.
– Quarks and leptons seem to fall into three distinct groups, or 'families'. No deep explanation is known.
– The mathematical consistency of the theory, known as 'the cancellation of the triangle anomalies', requires that the sum of all electric charges inside any family is equal to zero. This property has strong predictive power.

2.2 The electromagnetic interactions

All experimental data are well described by a simple Lagrangian interaction in which the photon field interacts with a current created from the fields of charged particles.

$$\mathcal{L}_i \sim eA_\mu(x)j^\mu(x) . \tag{1}$$

For the spinor matter fields of Table 1, the current takes the simple form

$$j^\mu(x) = \sum_i q_i \bar{\Psi}_i(x)\gamma^\mu\Psi_i(x) , \tag{2}$$

where q_i is the charge of the field Ψ_i in units of e.

This simple Lagrangian has some remarkable properties, all of which are verified by experiment.

- j is a vector current. The interaction separately conserves P, C and T.
- The current is diagonal in flavour space.
- More complex terms, such as $j^\mu(x)j_\mu(x)$ and $\partial A(x)\bar{\Psi}(x)\ldots\Psi(x),\ldots$ are absent, although they do not seem to be forbidden by any known property of the theory. All these terms, as well as all others we can write, share one common property: in a four-dimensional space–time, their canonical dimension is larger than four. We can easily show that the resulting quantum field theory is *non-renormalizable*. For some reason, nature does not like non-renormalizable theories.

Quantum electrodynamics (QED), the quantum field theory described by the Lagrangian in Eq. (1) and supplemented with the programme of renormalization, is one of the most successful physical theories. Its agreement with experiment is spectacular. For years it was the prototype for all other theories. The Standard Model is the result of the efforts to extend the ideas and methods of electromagnetic interactions to all other forces in physics.

2.3 The weak interactions

Weak interactions are mediated by massive vector bosons. When the Standard Model was proposed, their very existence as well as their number were unknown. But today we know that three massive vector bosons exist; two which are electrically charged and one which is neutral: W^+, W^- and Z^0. Like the photon, their couplings to matter are described by current operators:

$$\mathcal{L}_i \sim V_\mu(x)j^\mu(x); \quad V_\mu : \mathrm{W}^+_\mu, \mathrm{W}^-_\mu, \mathrm{Z}^0_\mu , \tag{3}$$

where the weak currents are again bi-linear in the fermion fields: $\bar{\Psi}\ldots\Psi$. Depending on the corresponding vector boson, we distinguish two types of weak currents: *the charged current*, coupled to W^+ and W^- and the *neutral current* coupled to Z^0, which have different properties.

The charged current:

- contains only left-handed fermion fields

$$j_\mu \sim \bar{\Psi}_\mathrm{L}\gamma_\mu\Psi_\mathrm{L} \sim \bar{\Psi}\gamma_\mu(1+\gamma_5)\Psi ; \tag{4}$$

- is non-diagonal in the quark flavour space;
- the coupling constants are complex.

The neutral current:

– contains both left- and right-handed fermion fields

$$j_\mu \sim C_\mathrm{L}\bar{\Psi}_\mathrm{L}\gamma_\mu\Psi_\mathrm{L} + C_\mathrm{R}\bar{\Psi}_\mathrm{R}\gamma_\mu\Psi_\mathrm{R}\ ; \tag{5}$$

– is diagonal in the quark flavour space.

With these currents, weak interactions have some properties which differ from those of the electromagnetic ones.

– Weak interactions violate P, C and T.
– In contrast to the photon, the weak vector bosons are self-coupled. The nature of these couplings is predicted theoretically in the framework of gauge theories and it has been determined experimentally.
– A new element has been added recently to the experimental landscape. It is a new scalar particle, compatible with what theorists have called *the Higgs boson*. Although all its properties have not yet been studied in detail, the existing evidence points towards the Higgs boson predicted by the Standard Model.

It is this kind of interaction that the Standard Model is supposed to describe.

3 Gauge symmetries

3.1 The concept of symmetry

In physics the concept of a symmetry follows from the assumption that a certain quantity is not measurable. As a result, the equations of motion should not depend on this quantity. We know from the general properties of classical mechanics that this implies the existence of conserved quantities. This relation between symmetries and conservation laws, epitomized by Noether's theorem, has been one of the most powerful tools in deciphering the properties of physical theories.

Some simple examples are given by the symmetries of space and time. The assumption that the position of the origin of the coordinate system is not physically measurable implies the invariance of the equations under space translations and the conservation of momentum. In the same way that we obtain the conservation laws of energy (time translations) and angular momentum (rotations), we can also distinguish between symmetries in *continuous transformations*, such as translations and rotations, and *discrete* symmetries, such as space or time inversions. Noether's theorem applies to the first. All symmetries of space and time are *geometrical* in the common sense of the word, and are easy to understand and visualize. During the last century we were led to consider two abstractions, each one of which has had a profound influence on our way of thinking about the fundamental interactions. Reversing the chronological order, we shall introduce first the idea of *internal* symmetries and second, that of local or *gauge* symmetries.

3.2 Internal symmetries

Internal symmetries are those with transformation parameters that do not affect the point of space and time x. The concept of such symmetries can be seen in classical physics, but it becomes natural in quantum mechanics and quantum field theory. The simplest example is the phase of the wave function. We know that it is not a measurable quantity, so the theory must be invariant under a change of phase. This is true for both relativistic or non-relativistic quantum mechanics. The equations of motion (Dirac or Schrödinger), as well as the normalization condition, are invariant under the transformation:

$$\Psi(x) \to \mathrm{e}^{\mathrm{i}\theta}\Psi(x)\ . \tag{6}$$

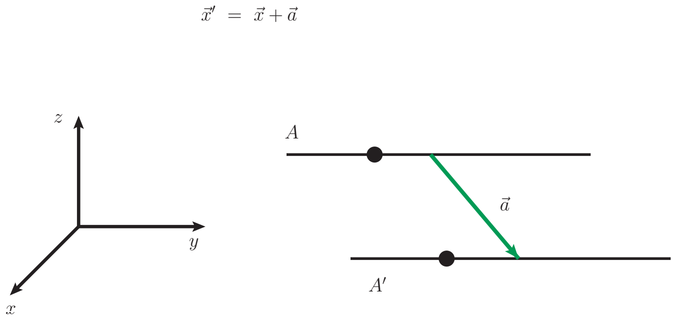

Fig. 1: A space translation by a constant vector $\vec{a}$

The transformation leaves the space–time point invariant, so it is an internal symmetry. Through Noether's theorem, invariance under Eq. (6) implies the conservation of the probability current.

The phase transformation in Eq. (6) corresponds to the Abelian group U(1). In 1932 Werner Heisenberg enlarged the concept to a non-Abelian symmetry with the introduction of isospin. The assumption is that strong interactions are invariant under a group of SU(2) transformations in which the proton and the neutron form a doublet $N(x)$:

$$N(x) = \begin{pmatrix} \mathrm{p}(x) \\ \\ \mathrm{n}(x) \end{pmatrix}; \quad N(x) \rightarrow \mathrm{e}^{\mathrm{i}\vec{\tau}\times\vec{\theta}} N(x) \,, \tag{7}$$

where $\vec{\tau}$ are proportional to the Pauli matrices and $\vec{\theta}$ are the three angles of a general rotation in a three-dimensional Euclidean space. Again, the transformations do not apply on the points of ordinary space.

Heisenberg's iso-space is three dimensional and isomorphic to our physical space. With the discovery of new internal symmetries the idea was generalized to multi-dimensional internal spaces. The space of physics, i.e. the space in which all symmetry transformations apply, became an abstract mathematical concept with non-trivial geometrical and topological properties. Only a part of it, the three-dimensional Euclidean space, is directly accessible to our senses.

3.3 Gauge symmetries

The concept of a local, or gauge, symmetry was introduced by Albert Einstein in his quest for the theory of general relativity[1]. Let us come back to the example of space translations, as shown in Fig. 1.

Figure 1 shows that if A is the trajectory of a free particle, then its image, after a translation of the form $\vec{x} \rightarrow \vec{x} + \vec{a}$, A', is also a possible trajectory of a free particle. The dynamics of free particles is invariant under space translations by a constant vector. It is a *global* invariance, in the sense that the parameter $\vec{a}$ is independent of the space–time point x. Is it possible to extend this invariance to a *local* one, namely one in which $\vec{a}$ is replaced by an arbitrary function of x; $\vec{a}(x)$? One usually calls the transformations in which the parameters are functions of the space–time point x *gauge transformations*[2] There may be various, essentially aesthetic, reasons for which one may wish to extend a global invariance to a gauge one. In physical terms, it can be argued that the formalism should allow for a local definition

[1]It is also present in classical electrodynamics if one considers the invariance under the change of the vector potential $A_\mu(x) \rightarrow A_\mu(x) - \partial_\mu\theta(x)$ with θ an arbitrary function, but before the introduction of quantum mechanics, this aspect of the symmetry was not emphasized.

[2]This strange terminology is due to Hermann Weyl. In 1918 he attempted to enlarge diffeomorphisms to local scale transformations and he called them, correctly, *gauge transformations*. The attempt was unsuccessful but, when he developed the theory for the Dirac electron in 1929, he still used the term gauge invariance, a term which has survived ever since, although the theory is no longer scale invariant.

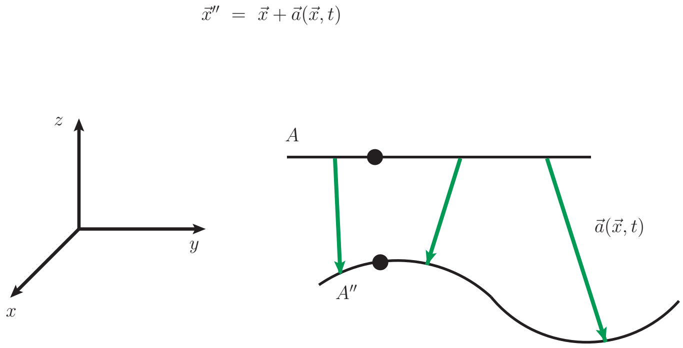

Fig. 2: A space translation by a vector $\vec{a}(x)$

of the origin of the coordinate system, since the latter is an unobservable quantity. From the mathematical point of view, local transformations produce a much richer and more interesting structure. Whichever one's motivations may be, physical or mathematical, it is clear that the free-particle dynamics is not invariant under translations in which $\vec{a}$ is replaced by $\vec{a}(x)$. This is shown schematically in Fig. 2.

We see that no free particle would follow the trajectory A''. This means that for A'' to be a trajectory, the particle must be subject to external forces. Can we determine these forces? The question sounds purely geometrical without any obvious physical meaning, so we expect a mathematical answer with no interest for physics. The great surprise is that the resulting theory, which is invariant under local translations, turns out to be classical general relativity, one of the four fundamental forces in nature. Gravitational interactions have such a geometric origin. In fact, the mathematical formulation of Einstein's original motivation to extend the principle of equivalence to accelerated frames is precisely the requirement of local invariance. Historically, many mathematical techniques which are used in today's gauge theories were developed in the framework of general relativity.

The gravitational forces are not the only ones that have a geometrical origin. Let us come back to the example of the quantum mechanical phase. It is clear that neither the Dirac nor the Schrödinger equation are invariant under a local change of phase $\theta(x)$. To be precise, let us consider the free Dirac Lagrangian,

$$\mathcal{L} = \bar{\Psi}(x)(\mathrm{i}\partial\!\!\!/ - m)\Psi(x) \; . \tag{8}$$

It is not invariant under the transformation

$$\Psi(x) \to \mathrm{e}^{\mathrm{i}\theta(x)}\Psi(x) \; . \tag{9}$$

The reason behind this is the presence of the derivative term in Eq. (8) which gives rise to a term proportional to $\partial_\mu\theta(x)$. In order to restore invariance, one must modify Eq. (8), in which case it will no longer describe a free Dirac field; invariance under gauge transformations leads to the introduction of interactions. Both physicists and mathematicians know the answer to the particular case of Eq. (8): one introduces a new field $A_\mu(x)$ and replaces the derivative operator ∂_μ by a 'covariant derivative' D_μ given by

$$D_\mu = \partial_\mu + \mathrm{i}eA_\mu \; , \tag{10}$$

where e is an arbitrary real constant. D_μ is said to be 'covariant' because it satisfies

$$D_\mu[\mathrm{e}^{\mathrm{i}\theta(x)}\Psi(x)] = \mathrm{e}^{\mathrm{i}\theta(x)}D_\mu\Psi(x) \; , \tag{11}$$

valid if, at the same time, $A_\mu(x)$ undergoes the transformation

$$A_\mu(x) \to A_\mu(x) - \frac{1}{e}\partial_\mu\theta(x) \; . \tag{12}$$

The Dirac Lagrangian density now becomes

$$\mathcal{L} = \bar{\Psi}(x)(\mathrm{i}\not{D} - m)\Psi(x) = \bar{\Psi}(x)(\mathrm{i}\not{\partial} - e\not{A} - m)\Psi(x) \; . \tag{13}$$

It is invariant under the gauge transformations of Eqs. (9) and (12) and describes the interaction of a charged spinor field with an external electromagnetic field! Replacing the derivative operator by the covariant derivative turns the Dirac equation into the same equation in the presence of an external electromagnetic field. Electromagnetic interactions give the same geometrical interpretation[3]. We can complete the picture by including the degrees of freedom of the electromagnetic field itself and add to Eq. (13) the corresponding Lagrangian density. Again, gauge invariance determines its form uniquely and we are led to the well-known result

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \bar{\Psi}(x)(\mathrm{i}\not{D} - m)\Psi(x) \tag{14}$$

with

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) \; . \tag{15}$$

The constant e we introduced is the electric charge, the coupling strength of the field Ψ with the electromagnetic field. Notice that a second field Ψ' will be coupled with its own charge e'.

Let us summarize: we started with a theory invariant under a group U(1) of global phase transformations. The extension to a local invariance can be interpreted as a U(1) symmetry at each point x. In a qualitative way we can say that gauge invariance induces an invariance under $\mathrm{U}(1)^\infty$. We saw that this extension, a purely geometrical requirement, implies the introduction of new interactions. The surprising result here is that these 'geometrical' interactions describe the well-known electromagnetic forces.

The extension of the formalism of gauge theories to non-Abelian groups is not trivial and was first discovered by trial and error. Here we shall restrict ourselves to internal symmetries which are simpler to analyse and they are the ones we shall apply to particle physics outside gravitation.

Let us consider a classical field theory given by a Lagrangian density $\mathcal{L}$. It depends on a set of N fields $\psi^i(x)$, $i = 1, \dots, r$, and their first derivatives. The Lorentz transformation properties of these fields will play no role in this discussion. We assume that the ψ transform linearly according to an r-dimensional representation, not necessarily irreducible, of a compact, simple Lie group, G, which does not act on the space–time point x.

$$\Psi = \begin{pmatrix} \psi^1 \\ \vdots \\ \psi^r \end{pmatrix}, \quad \Psi(x) \to U(\omega)\Psi(x), \quad \omega \in G \; , \tag{16}$$

where $U(\omega)$ is the matrix of the representation of G. In fact, in these lectures we shall be dealing only with perturbation theory and it will be sufficient to look at transformations close to the identity in G.

$$\Psi(x) \to \mathrm{e}^{\mathrm{i}\Theta}\Psi(x), \quad \Theta = \sum_{a=1}^{m} \theta^a T^a \tag{17}$$

[3]The same applies to the Schrödinger equation. In fact, this was done first by V. Fock in 1926, immediately after Schrödinger's original publication.

where the θ^a are a set of m constant parameters, and the T^a are m $r \times r$ matrices representing the m generators of the Lie algebra of G. They satisfy the commutation rules

$$[T^a, T^b] = \mathrm{i} f^{abc} T^c \ . \tag{18}$$

The f are the structure constants of G and a summation over repeated indices is understood. The normalization of the structure constants is usually fixed by requiring that, in the fundamental representation, the corresponding matrices of the generators t^a are normalized such as

$$\mathrm{Tr}(t^a t^b) = \frac{1}{2}\delta^{ab} \ . \tag{19}$$

The Lagrangian density $\mathcal{L}(\Psi, \partial\Psi)$ is assumed to be invariant under the global transformations of Eqs. (16) or (17). As was done for the Abelian case, we wish to find a new $\mathcal{L}$, invariant under the corresponding gauge transformations in which the θ^a of Eq. (17) are arbitrary functions of x. In the same qualitative sense, we look for a theory invariant under G^∞. This problem, stated the way we present it here, was first solved by trial and error for the case of SU(2) by C.N. Yang and R.L. Mills in 1954. They gave the underlying physical motivation and these theories are called since 'Yang–Mills theories'. The steps are direct generalizations of the ones followed in the Abelian case. We need a gauge field, the analogue of the electromagnetic field, to transport the information contained in Eq. (17) from point to point. Since we can perform m independent transformations, the number of generators in the Lie algebra of G, we need m gauge fields $A^a_\mu(x)$, $a = 1, \ldots, m$. It is easy to show that they belong to the adjoint representation of G. Using the matrix representation of the generators we can cast $A^a_\mu(x)$ into an $r \times r$ matrix:

$$\mathcal{A}_\mu(x) = \sum_{a=1}^{m} A^a_\mu(x) T^a \ . \tag{20}$$

The covariant derivatives can now be constructed as

$$\mathcal{D}_\mu = \partial_\mu + \mathrm{i}g\mathcal{A}_\mu \ , \tag{21}$$

with g as an arbitrary real constant. They satisfy

$$\mathcal{D}_\mu \mathrm{e}^{\mathrm{i}\Theta(x)}\Psi(x) = \mathrm{e}^{\mathrm{i}\Theta(x)}\mathcal{D}_\mu\Psi(x) \ , \tag{22}$$

provided the gauge fields transform as

$$\mathcal{A}_\mu(x) \to \mathrm{e}^{\mathrm{i}\Theta(x)}\mathcal{A}_\mu(x)\mathrm{e}^{-\mathrm{i}\Theta(x)} + \frac{\mathrm{i}}{g}\left(\partial_\mu \mathrm{e}^{\mathrm{i}\Theta(x)}\right)\mathrm{e}^{-\mathrm{i}\Theta(x)} \ . \tag{23}$$

The Lagrangian density $\mathcal{L}(\Psi, \mathcal{D}\Psi)$ is invariant under the gauge transformations of Eqs. (17) and (23) with an x-dependent Θ, if $\mathcal{L}(\Psi, \partial\Psi)$ is invariant under the corresponding global ones of Eqs. (16) or (17). As with the electromagnetic field, we can include the degrees of freedom of the new gauge fields by adding to the Lagrangian density a gauge invariant kinetic term. It turns out that it is slightly more complicated than $F_{\mu\nu}$ of the Abelian case. Yang and Mills computed it for SU(2) but it is uniquely determined by geometry plus some obvious requirements, such as absence of higher-order derivatives. The result is given by

$$\mathcal{G}_{\mu\nu} = \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu - \mathrm{i}g\left[\mathcal{A}_\mu, \mathcal{A}_\nu\right] \ . \tag{24}$$

The full gauge-invariant Lagrangian can now be written as

$$\mathcal{L}_{\rm inv} = -\frac{1}{2} Tr \mathcal{G}_{\mu\nu}\mathcal{G}^{\mu\nu} + \mathcal{L}(\Psi, \mathcal{D}\Psi) \; . \tag{25}$$

By convention, in Eq. (24) the matrix $\mathcal{A}$ is taken to be

$$\mathcal{A}_\mu = A^a_\mu t^a \; , \tag{26}$$

where we recall that the t^a are the matrices representing the generators in the fundamental representation. It is only with this convention that the kinetic term in Eq. (25) is correctly normalized. In terms of the component fields A^a_μ, $\mathcal{G}_{\mu\nu}$ reads

$$\mathcal{G}_{\mu\nu} = G^a_{\mu\nu} t^a, \quad G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu \; . \tag{27}$$

Under a gauge transformation $\mathcal{G}_{\mu\nu}$ transforms like a member of the adjoint representation:

$$\mathcal{G}_{\mu\nu}(x) \to \mathrm{e}^{\mathrm{i}\theta^a(x)t^a} \;\; \mathcal{G}_{\mu\nu}(x) \;\; \mathrm{e}^{-\mathrm{i}\theta^a(x)t^a} \; . \tag{28}$$

This completes the construction of the gauge invariant Lagrangian. We add some remarks below.

– As was the case with the electromagnetic field, the Lagrangian of Eq. (25) does not contain terms proportional to $A_\mu A^\mu$. This means that, under the usual quantization rules, the gauge fields describe massless particles.
– Since $\mathcal{G}_{\mu\nu}$ is not linear in the fields $\mathcal{A}_\mu$, the $\mathcal{G}^2$ term in Eq. (25), besides the usual kinetic term which is bilinear in the fields, contains tri-linear and quadri-linear terms. In perturbation theory, they will be treated as coupling terms whose strength is given by the coupling constant g. In other words, the non-Abelian gauge fields are self-coupled while the Abelian (photon) field is not. A Yang–Mills theory, containing only gauge fields, is still a dynamically rich quantum field theory, whereas a theory with the electromagnetic field alone is a trivial free theory.
– The same coupling constant g appears in the covariant derivative of the fields Ψ in Eq. (21). This simple consequence of gauge invariance has an important physical application: if we add another field Ψ', its coupling strength with the gauge fields will still be given by the same constant g. Contrary to the Abelian case studied before, if electromagnetism is part of a non-Abelian simple group, gauge invariance implies charge quantization.
– The above analysis can be extended in a straightforward way to the case where the group G is the product of simple groups $G = G_1 \times \cdots \times G_n$. The only difference is that one should introduce n coupling constants $g_1, \ldots, g_n$, one for each simple factor. Charge quantization is still true inside each subgroup, but charges belonging to different factors are no longer related.
– The situation changes if one considers non semi-simple groups, where one or more of the factors G_i is Abelian. In this case, the associated coupling constants can be chosen different for each field and the corresponding Abelian charges are not quantized.

As we alluded to above, gauge theories have a deep geometrical meaning. In order to get a better understanding of this property without entering into complicated issues of differential geometry, it is instructive to consider a reformulation of the theory replacing the continuum of space–time with a four-dimensional Euclidean lattice. We can do that very easily. Let us consider, for simplicity, a lattice with hypercubic symmetry. The space–time point x_μ is replaced by

$$x_\mu \to n_\mu a \; , \tag{29}$$

where a is a constant length (the lattice spacing) and n_μ is a d-dimensional vector with components $n_\mu = (n_1, n_2, \ldots, n_d)$ which take integer values $0 \leq n_\mu \leq N_\mu$. N_μ is the number of points of our lattice in the direction μ. The total number of points, i.e. the volume of the system, is given by $V \sim \prod_{\mu=1}^{d} N_\mu$. The presence of a introduces an ultraviolet, or short distance, cut-off because all momenta are bounded from above by $2\pi/a$. The presence of N_μ introduces an infrared or large distance cut-off because the momenta are also bounded from below by $2\pi/Na$, where N is the maximum of N_μ. The infinite volume–continuum space is recovered at the double limit $a \to 0$ and $N_\mu \to \infty$.

The dictionary between quantities defined in the continuum and the corresponding ones on the lattice is easy to establish (we take the lattice spacing a equal to one):

- a field $\Psi(x) \quad \Rightarrow \quad \Psi_n$,
  where the field Ψ is an r-component column vector as in Eq. (16);
- a local term such as $\bar{\Psi}(x)\Psi(x) \quad \Rightarrow \quad \bar{\Psi}_n\Psi_n$;
- a derivative $\partial_\mu\Psi(x) \quad \Rightarrow \quad (\Psi_n - \Psi_{n+\mu})$,
  where $n+\mu$ should be understood as a unit vector joining the point n with its nearest neighbour in the direction μ;
- the kinetic energy term[4] $\bar{\Psi}(x)\partial_\mu\Psi(x) \quad \Rightarrow \quad \bar{\Psi}_n\Psi_n - \bar{\Psi}_n\Psi_{n+\mu}$.

We may be tempted to write similar expressions for the gauge fields, but we must be careful with the way gauge transformations act on the lattice. Let us repeat the steps we followed in the continuum. Under gauge transformations a field transforms as:

- gauge transformations $\Psi(x) \to \mathrm{e}^{\mathrm{i}\Theta(x)}\Psi(x) \quad \Rightarrow \quad \Psi_n \to \mathrm{e}^{\mathrm{i}\Theta_n}\Psi_n$,
  so all local terms of the form $\bar{\Psi}_n\Psi_n$ remain invariant but the part of the kinetic energy which couples fields at neighbouring points does not;
- the kinetic energy $\bar{\Psi}_n\Psi_{n+\mu} \to \bar{\Psi}_n\mathrm{e}^{-\mathrm{i}\Theta_n}\mathrm{e}^{\mathrm{i}\Theta_{n+\mu}}\Psi_{n+\mu}$,
  which shows that we recover the problem we had with the derivative operator in the continuum.

In order to restore invariance we must introduce a new field, which is an $r\times r$ matrix, and which has indices n and $n+\mu$. We denote it by $U_{n,n+\mu}$ and we shall impose on it the constraint $U_{n,n+\mu} = U^{-1}_{n+\mu,n}$. Under a gauge transformation, U transforms as

$$U_{n,n+\mu} \quad \to \quad \mathrm{e}^{\mathrm{i}\Theta_n}U_{n,n+\mu}\mathrm{e}^{-\mathrm{i}\Theta_{n+\mu}} \ . \tag{30}$$

With the help of this gauge field we write the kinetic-energy term with the covariant derivative on the lattice as:

$$\bar{\Psi}_n\, U_{n,n+\mu}\, \Psi_{n+\mu} \ , \tag{31}$$

which is invariant under gauge transformations.

U is an element of the gauge group but we can show that, at the continuum limit and for an infinitesimal transformation, it correctly reproduces A_μ, which belongs to the Lie algebra of the group. Notice that, contrary to the field Ψ, U does not live on a single lattice point, but it has two indices, n and $n+\mu$, in other words it lives on the oriented link joining the two neighbouring points. We see here that the mathematicians are right when they do not call the gauge field 'a field' but 'a connection'.

In order to finish the story we want to obtain an expression for the kinetic energy of the gauge field, the analogue of $Tr\mathcal{G}_{\mu\nu}(x)\mathcal{G}^{\mu\nu}(x)$, on the lattice. As for the continuum, the guiding principle is gauge invariance. Let us consider two points on the lattice n and m. We shall call a path $p_{n,m}$ on the lattice a sequence of oriented links which continuously join the two points. Next, consider the product of the gauge fields U along all the links of the path $p_{n,m}$:

[4] We write here the expression for spinor fields which contain only first-order derivatives in the kinetic energy. The extension to scalar fields with second-order derivatives is obvious.

$$P^{(p)}(n,m) = \prod_p U_{n,n+\mu} \cdots U_{m-\nu,m} \,. \tag{32}$$

Using the transformation rule in Eq. (30), we see that $P^{(p)}(n,m)$ transforms as

$$P^{(p)}(n,m) \to \mathrm{e}^{\mathrm{i}\Theta_n} P^{(p)}(n,m) \mathrm{e}^{-\mathrm{i}\Theta_m} \,. \tag{33}$$

It follows that if we consider a closed path $c = p_{n,n}$, the quantity Tr $P^{(c)}$ is gauge invariant. The simplest closed path for a hypercubic lattice has four links and it is called a *plaquette.* The correct form of the Yang–Mills action on the lattice can be written in terms of the sum of Tr $P^{(c)}$ over all plaquettes.

4 Spontaneous symmetry breaking

Since gauge theories appear to predict the existence of massless gauge bosons, when they were first proposed they did not seem to have any direct application to particle physics outside electromagnetism. It is this handicap which plagued gauge theories for many years. In this section, we shall present a seemingly unrelated phenomenon that will turn out to provide the answer.

An infinite system may exhibit the phenomenon of phase transitions. It often implies a reduction in the symmetry of the ground state. A field theory is a system with an infinite number of degrees of freedom, so it is not surprising that field theories may also show the phenomenon of phase transitions. Let us consider the example of a field theory invariant under a set of transformations forming a group G. In many cases, we encounter at least two phases.

– *The unbroken* or *the Wigner phase*: the symmetry is manifest in the spectrum of the theory whose excitations form irreducible representations of the symmetry group. For a gauge theory, the vector gauge bosons are massless and belong to the adjoint representation. But we have good reason to believe that, for non-Abelian gauge theories, a strange phenomenon occurs in this phase: all physical states are singlets of the group. All non-singlet states, such as those corresponding to the gauge fields, are supposed to be *confined*, in the sense that they do not appear as physically realizable asymptotic states.
– *The spontaneously broken phase*: part of the symmetry is hidden from the spectrum. For a gauge theory, some of the gauge bosons become massive and appear as physical states.

It is this kind of phase transition that we want to study in this section.

4.1 An example from classical mechanics

A very simple example is provided by the problem of the bent rod. Let a cylindrical rod be charged as in Fig. 3. The problem is obviously symmetric under rotations around the z-axis. Let z measure the distance from the basis of the rod, and $X(z)$ and $Y(z)$ give the deviations, along the x and y directions respectively, of the axis of the rod at the point z from the symmetric position. For small deflections the equations of elasticity can be linearized and take the form

$$IE\frac{\mathrm{d}^4X}{\mathrm{d}z^4} + F\frac{\mathrm{d}^2X}{\mathrm{d}z^2} = 0\,; \quad IE\frac{\mathrm{d}^4Y}{\mathrm{d}z^4} + F\frac{\mathrm{d}^2Y}{\mathrm{d}z^2} = 0\,. \tag{34}$$

where $I = \pi R^4/4$ is the moment of inertia of the rod and E is the Young modulus. It is obvious that the system shown in Eq. (34) always possesses a symmetric solution $X = Y = 0$. However, we can also look for asymmetric solutions of the general form $X = A + Bz + C\sin kz + D\cos kz$ with $k^2 = F/EI$, which satisfy the boundary conditions $X = X'' = 0$ at $z = 0$ and $z = l$. We find that such solutions

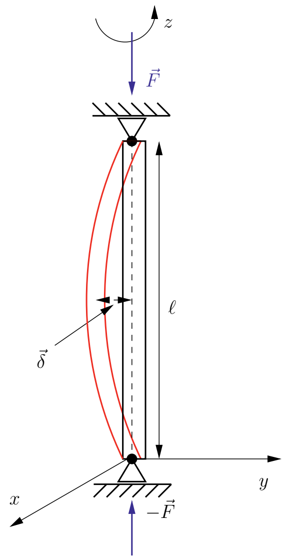

Fig. 3: A cylindrical rod bent under a force F along its symmetry axis

exist, $X = C \sin kz$, provided $kl = n\pi; n = 1, \ldots$. The first such solution appears when F reaches a critical value $F_{\rm cr}$ given by

$$F_{\rm cr} = \frac{\pi^2 EI}{l^2} \, . \tag{35}$$

The appearance of these solutions is already an indication of instability and a careful study of the stability problem proves that the non-symmetric solutions correspond to lower energy. From that point Eq. (34) is no longer valid because they only apply to small deflections, and we must use the general equations of elasticity. The result is that this instability of the symmetric solution occurs for all values of F larger than $F_{\rm cr}$

What has happened to the original symmetry of the equations? It is still hidden in the sense that we cannot predict in which direction the rod is going to bend in the x–y plane. They all correspond to solutions with precisely the same energy. In other words, if we apply a symmetry transformation (in this case a rotation around the z-axis) to an asymmetric solution, we obtain another asymmetric solution which is degenerate with the first one.

We call such a symmetry 'spontaneously broken', and in this simple example we see all its characteristics:

- there exists a critical point, i.e., a critical value of some external quantity which we can vary freely (in this case the external force F; in several physical systems it is the temperature) which determines whether spontaneous symmetry breaking will take place or not. Beyond this critical point:
  - the symmetric solution becomes unstable;
  - the ground state becomes degenerate.

The complete mathematical analysis of this system requires the study of the exact equations of elasticity which are non-linear, but we can look at a simplified version. A quantity, which plays an important role in every phenomenon of phase transition, is *the order parameter*, whose value determines in which phase the system is. In our example, we choose it to be the two-component vector $\vec{\delta}$ shown in Fig. 3, which we write as a complex number $\delta = \delta_x + {\rm i}\delta_y$ with $\delta = \rho {\rm e}^{{\rm i}\theta}$. The symmetric phase corresponds to $\rho = 0$. It is instructive to express the energy of the system E as a function of the order parameter. Rotational invariance implies that E depends only on $\vec{\delta} \times \vec{\delta} = \rho^2$. At the vicinity of the critical point ρ^2 is small and we can expand E as

$$E = C_0 + C_1\rho^2 + C_2\rho^4 + \cdots \ . \tag{36}$$

The C are constants which depend on the characteristics of the rod and the force F. Stability is obtained by

$$\frac{\mathrm{d}E}{\mathrm{d}\rho}(\rho = v) = 0 \ \Rightarrow \ v(C_1 + 2C_2v^2) = 0 \ . \tag{37}$$

We thus find the two solutions we mentioned above, namely $v = 0$ for the symmetric case and $v^2 = -C_1/2C_2$ for the spontaneously broken phase. Since ρ is real, this second solution is acceptable if C_1/C_2 is negative. C_2 must be positive for the energy to be bounded from below in the approximate Eq. (36). Therefore, C_1 must vanish at the critical point and change sign with $F - F_{\mathrm{cr}}$. As a result, we can write $C_1 = \hat{C}_1(F_{\mathrm{cr}} - F)$ with $\hat{C}_1 > 0$. For $F > F_{\mathrm{cr}}$, C_1 is negative and we can write the energy as

$$E = C_0 + \hat{C}_1(F_{\mathrm{cr}} - F)\vec{\delta} \times \vec{\delta} + C_2(\vec{\delta} \times \vec{\delta})^2 = \hat{C}_1(F - F_{\mathrm{cr}})\frac{(\rho^2 - v^2)^2}{2v^2} \ , \tag{38}$$

with v given by the non-zero solution of Eq. (37). With the energy defined up to an arbitrary additive constant, we have fixed C_0 by the condition that the energy of the ground state $\rho = v$ vanishes. In the phase with spontaneous symmetry breaking, the energy of the symmetric $\rho = 0$ solution is positive and given by

$$E_0 = \hat{C}_1(F - F_{\mathrm{cr}})\frac{v^2}{2} \ . \tag{39}$$

The expression for the energy given by Eq. (38) has the well-known form of Fig. 4 with a single minimum $v = 0$ for $F < F_{\mathrm{cr}}$ and the Mexican hat form for $F > F_{\mathrm{cr}}$.

There are a great variety of physical systems, both in classical and quantum physics, exhibiting spontaneous symmetry breaking, but we will not describe any others here. The Heisenberg ferro-magnet is a good example to keep in mind, because we shall often use it as a guide, but no essentially new phenomenon appears outside the ones already described. Therefore, we shall go directly to some field theory models.

4.2 A simple field theory model

Let $\phi(x)$ be a complex scalar field whose dynamics is described by the Lagrangian density

$$\mathcal{L}_1 \ = \ (\partial_\mu\phi)(\partial^\mu\phi^*) - M^2\phi\phi^* - \lambda(\phi\phi^*)^2 \ , \tag{40}$$

where $\mathcal{L}_1$ is a classical Lagrangian density and $\phi(x)$ is a classical field. No quantization is considered for the moment. Eq. (40) is invariant under the group U(1) of global transformations:

$$\phi(x) \ \rightarrow \ \mathrm{e}^{\mathrm{i}\theta}\phi(x) \ . \tag{41}$$

The current, $j_\mu \sim \phi\partial_\mu\phi^* - \phi^*\partial_\mu\phi$, whose conservation can be verified using the equations of motion, corresponds to this invariance.

We are interested in the classical field configuration which minimizes the energy of the system. We thus compute the Hamiltonian density given by

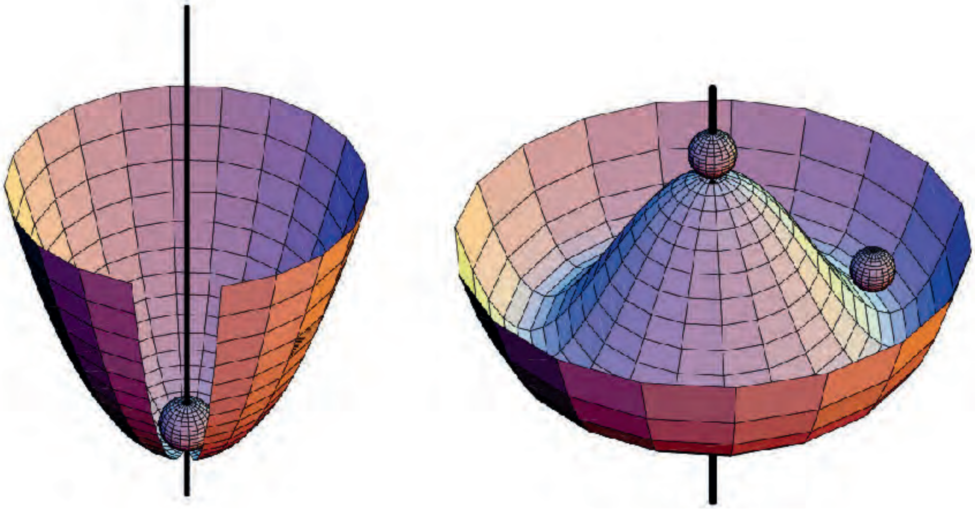

Fig. 4: The potential $V(\phi)$ with $M^2 \geq 0$ (left) and M^2 <0 (right)

$$\mathcal{H}_1 = (\partial_0\phi)(\partial_0\phi^*) + (\partial_i\phi)(\partial_i\phi^*) + V(\phi) , \tag{42}$$

$$V(\phi) = M^2\phi\phi^* + \lambda(\phi\phi^*)^2 . \tag{43}$$

The first two terms of $\mathcal{H}_1$ are positive definite. They can only vanish for ϕ = constant. Therefore, the ground state of the system corresponds to ϕ = constant = minimum of $V(\phi)$. V has a minimum only if $\lambda > 0$. In this case, the position of the minimum depends on the sign of M^2. (Notice that we are still studying a classical field theory and M^2 is just a parameter. One should not be misled by the notation into thinking that M is a 'mass' and M^2 is necessarily positive.)

For $M^2 > 0$, the minimum is at ϕ = 0 (symmetric solution, shown in the left-hand side of Fig. 4), but for $M^2 < 0$ there is a whole circle of minima at the complex ϕ-plane with radius $v = (-M^2/2\lambda)^{1/2}$ (Fig. 4, right-hand side). Any point on the circle corresponds to a spontaneous breaking of Eq. (41).

We see that:

- the critical point is M^2 = 0;
- for $M^2 > 0$ the symmetric solution is stable;
- for $M^2 < 0$ spontaneous symmetry breaking occurs.

Let us choose $M^2 < 0$. In order to reach the stable solution we translate the field ϕ. It is clear that there is no loss of generality by choosing a particular point on the circle, since they are all obtained from any given one by applying the transformations from Eq. (41). Let us, for convenience, choose the point on the real axis in the ϕ-plane. We thus write

$$\phi(x) = \frac{1}{\sqrt{2}}\left[v + \psi(x) + \mathrm{i}\chi(x)\right] . \tag{44}$$

Bringing (44) in (40) we find

$$\begin{aligned}\mathcal{L}_1(\phi) \;\rightarrow\; \mathcal{L}_2(\psi,\chi) &= \frac{1}{2}(\partial_\mu\psi)^2 + \frac{1}{2}(\partial_\mu\chi)^2 - \frac{1}{2}(2\lambda v^2)\psi^2 \\ &\quad - \lambda v\psi(\psi^2+\chi^2) - \frac{\lambda}{4}(\psi^2+\chi^2)^2 .\end{aligned} \tag{45}$$

Notice that $\mathcal{L}_2$ does not contain any term proportional to χ^2, which is expected since V is locally flat in the χ direction. A second remark concerns the arbitrary parameters of the theory. $\mathcal{L}_1$ contains two such parameters: M, which has the dimensions of a mass, and λ, a dimensionless coupling constant. In $\mathcal{L}_2$ we again have the coupling constant λ and a new mass parameter v which is a function of M and λ. It is important to notice that, although $\mathcal{L}_2$ also contains trilinear terms, its coupling strength is not a new parameter but is proportional to $v\lambda$. $\mathcal{L}_2$ is still invariant under the transformations with infinitesimal parameter θ:

$$\delta\psi = -\theta\chi \; ; \qquad \delta\chi = \theta\psi + \theta v \; , \tag{46}$$

to which corresponds a conserved current

$$j_\mu \sim \psi\partial_\mu\chi - \chi\partial_\mu\psi + v\partial_\mu\chi \; . \tag{47}$$

The last term, which is linear in the derivative of χ, is characteristic of the phenomenon of spontaneous symmetry breaking.

It should be emphasized here that $\mathcal{L}_1$ and $\mathcal{L}_2$ are completely equivalent Lagrangians. They both describe the dynamics of the same physical system and a change of variables, as in Eq. (44), cannot change the physics. However, this equivalence is only true if we can solve the problem exactly. In this case, we shall find the same solution using either of them. However, we do not have exact solutions and we intend to apply perturbation theory, which is an approximation scheme. The equivalence is then no longer guaranteed and perturbation theory has much better chances to give sensible results using one language rather than the other. In particular, if we use $\mathcal{L}_1$ as a quantum field theory and we decide to apply perturbation theory, using the quadratic terms of $\mathcal{L}_1$ as the unperturbed part, we immediately see that we shall get nonsense. The spectrum of the unperturbed Hamiltonian would consist of particles with negative square mass, and no perturbation corrections at any finite order could change that. This is essentially because we are trying to calculate the quantum fluctuations around an unstable solution and perturbation theory is just not designed to do that. On the contrary, we see that the quadratic part of $\mathcal{L}_2$ gives a reasonable spectrum; thus we hope that perturbation theory will also give reasonable results. Therefore, we conclude that our physical system, considered now as a quantum system, consists of two interacting scalar particles, one with mass $m_\psi^2 = 2\lambda v^2$ and the other with $m_\chi = 0$. We believe that this is the spectrum we would have also found starting from $\mathcal{L}_1$, if we could solve the dynamics exactly.

The appearance of a zero-mass particle in the quantum version of the model is an example of a general theorem attributable to J. Goldstone: for every generator of a spontaneously broken symmetry there corresponds a massless particle, called the Goldstone particle. This theorem is just the translation of the statement about the degeneracy of the ground state into quantum-field-theory language. The ground state of a system described by a quantum field theory is the vacuum state, and you need massless excitations in the spectrum of states in order to allow for the degeneracy of the vacuum.

4.3 Gauge symmetries

In this section, we want to study the consequences of spontaneous symmetry breaking in the presence of a gauge symmetry. We shall find a very surprising result. When combined together, the two problems, namely the massless gauge bosons on the one hand and the massless Goldstone bosons on the other, will solve each other. It is this miracle that we want to present here[5]. We start with the Abelian case.

We look at the model of the previous section in which the U(1) symmetry of Eq. (41) has been promoted to a local symmetry with $\theta \to \theta(x)$. As we explained already, this implies the introduction of

[5]In relativistic physics this mechanism was invented and developed by François Englert and Robert Brout, Peter Higgs, as well as Gerald Guralnik, Carl Richard Hagen and Thomas Walter Bannerman Kibble.

a massless vector field, which we can call the 'photon' and the interactions are obtained by replacing the derivative operator ∂_μ by the covariant derivative D_μ and adding the photon kinetic energy term:

$$\mathcal{L}_1 = -\frac{1}{4}F_{\mu\nu}^2 + |(\partial_\mu + \mathrm{i}eA_\mu)\phi|^2 - M^2\phi\phi^* - \lambda(\phi\phi^*)^2 \; . \tag{48}$$

$\mathcal{L}_1$ is invariant under the gauge transformation:

$$\phi(x) \;\;\rightarrow\;\; \mathrm{e}^{\mathrm{i}\theta(x)}\phi(x)\; ; \quad A_\mu \;\;\rightarrow\;\; A_\mu - \frac{1}{e}\partial_\mu\theta(x)\; . \tag{49}$$

The same analysis as before shows that for $\lambda > 0$ and $M^2 < 0$ there is a spontaneous breaking of the U(1) symmetry. Replacing Eq. (44) for (48) we obtain

$$\begin{aligned}\mathcal{L}_1 \;\rightarrow\; \mathcal{L}_2 = &-\frac{1}{4}F_{\mu\nu}^2 + \frac{e^2v^2}{2}A_\mu^2 + evA_\mu\partial^\mu\chi \\ &+\frac{1}{2}(\partial_\mu\psi)^2 + \frac{1}{2}(\partial_\mu\chi)^2 - \frac{1}{2}(2\lambda v^2)\psi^2 \;\; + \cdots \; ,\end{aligned} \tag{50}$$

where the dots stand for coupling terms which are at least trilinear in the fields.

The surprising term is the second one, which is proportional to A_μ^2. It looks as though the photon has become massive. Notice that Eq. (50) is still gauge invariant since it is equivalent to Eq. (48). The gauge transformation is now obtained by replacing Eq. (44) with Eq. (49):

$$\begin{aligned}\psi(x) \;&\rightarrow\; \cos\theta(x)[\psi(x)+v] - \sin\theta(x)\chi(x) - v \\ \chi(x) \;&\rightarrow\; \cos\theta(x)\chi(x) + \sin\theta(x)[\psi(x)+v] \\ A_\mu \;&\rightarrow\; A_\mu - \frac{1}{e}\partial_\mu\theta(x)\; .\end{aligned} \tag{51}$$

This means that our previous conclusion, that gauge invariance forbids the presence of an A_μ^2 term, was simply wrong. Such a term can be present, but the gauge transformation is slightly more complicated; it must be accompanied by a translation of the field.

The Lagrangian of Eq. (50), if taken as a quantum field theory, seems to describe the interaction of a massive vector particle (A_μ) and two scalars, one massive (ψ) and one massless (χ). However, we can immediately see that something is wrong with this counting. A warning is already contained in the non-diagonal term between A_μ and $\partial^\mu\chi$. Indeed, the perturbative particle spectrum can be read from the Lagrangian only after we have diagonalized the quadratic part. A more direct way to see the trouble is to count the apparent degrees of freedom[6] before and after the translation:

– Lagrangian of Eq. (48):
  (i) one massless vector field: 2 degrees;
  (ii) one complex scalar field: 2 degrees;
  total: 4 degrees.

[6] The terminology here is misleading. As we pointed out earlier, any field theory, considered as a dynamical system, is a system with an infinite number of degrees of freedom. For example, the quantum theory of a free neutral scalar field is described by an infinite number of harmonic oscillators, one for every value of the three-dimentional momentum. Here, we use the same term 'degrees of freedom' to denote the independent one-particle states. We know that for a massive spin-s particle we have $2s+1$ one-particle states, and for a massless particle with spin other than zero we only have two. In fact, it would have been more appropriate to talk about a $(2s+1)$-*infinity* and 2-*infinity* degrees of freedom, respectively.

– Lagrangian of Eq. (50):
  (i) one massive vector field: 3 degrees;
  (ii) two real scalar fields: 2 degrees;
  total: 5 degrees.

Since physical degrees of freedom cannot be created by a simple change of variables, we conclude that the Lagrangian of Eq. (50) must contain fields which do not create physical particles. This is indeed the case, and we can show a transformation which makes the unphysical fields disappear. Instead of parametrizing the complex field ϕ by its real and imaginary parts, let us choose its modulus and its phase. The choice is dictated by the fact that it is a change of phase that describes the motion along the circle of the minima of the potential $V(\phi)$. We thus write

$$\phi(x) = \frac{1}{\sqrt{2}}[v+\rho(x)]\mathrm{e}^{\mathrm{i}\zeta(x)/v} \; ; \quad A_\mu(x) = B_\mu(x) - \frac{1}{ev}\partial_\mu\zeta(x) \; . \tag{52}$$

In this notation, the gauge transformation Eq. (49) or Eq. (51) is simply a translation of the field ζ: $\zeta(x) \to \zeta(x) + v\theta(x)$. Replacing Eq. (52) with Eq. (48) we obtain

$$\begin{aligned} \mathcal{L}_1 \;\to\; \mathcal{L}_3 &= -\frac{1}{4}B_{\mu\nu}^2 + \frac{e^2v^2}{2}B_\mu^2 + \frac{1}{2}(\partial_\mu\rho)^2 - \frac{1}{2}(2\lambda v^2)\rho^2 \\ &\quad - \frac{\lambda}{4}\rho^4 + \frac{1}{2}e^2B_\mu^2(2v\rho+\rho^2) \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \; . \end{aligned} \tag{53}$$

The $\zeta(x)$ field has disappeared. Equation (53) describes two massive particles, a vector (B_μ) and a scalar (ρ). It exhibits no gauge invariance, since the original symmetry $\zeta(x) \to \zeta(x) + v\theta(x)$ is now trivial.

We see that there are three different Lagrangians describing the same physical system. $\mathcal{L}_1$ is invariant under the usual gauge transformation, but it contains a negative square mass and it is therefore unsuitable for quantization. $\mathcal{L}_2$ is still gauge invariant, but the transformation law from Eq. (51) is more complicated. It can be quantized in a space containing unphysical degrees of freedom. This by itself is not a great obstacle and it occurs frequently. For example, ordinary QED is usually quantized in a space involving unphysical (longitudinal and scalar) photons. In fact, it is $\mathcal{L}_2$, in a suitable gauge, which is used for general proofs of renormalizability as well as for practical calculations. Finally, $\mathcal{L}_3$ is no longer invariant under any kind of gauge transformation, but clearly exhibits the particle spectrum of the theory. It contains only physical particles and they are all massive. This is the miracle that was announced earlier. Although we start from a gauge theory, the final spectrum contains massive particles only. Actually, $\mathcal{L}_3$ can be obtained from $\mathcal{L}_2$ by an appropriate choice of gauge. The conclusion so far can be stated as follows.

In a spontaneously broken gauge theory, the gauge vector bosons acquire a mass and the would-be massless Goldstone bosons decouple and disappear. Their degrees of freedom are used to make the transition from massless to massive vector bosons possible.

The extension to the non-Abelian case is straightforward. Let us consider a gauge group G with m generators and, thus, m massless gauge bosons. The claim is that we can break part of the symmetry spontaneously, leaving a subgroup H with h generators unbroken. The h gauge bosons associated with H remain massless while the $m-h$ others acquire a mass. In order to achieve this result we need $m-h$ scalar degrees of freedom with the same quantum numbers as the broken generators. They will disappear from the physical spectrum and will re-appear as zero-helicity states of the massive vector bosons. As previously, we shall see that one needs at least one more scalar state which remains physical.

In the remaining part of this section, we show these results for a general gauge group. The reader who is not interested in technical details may skip this part.

We introduce a multiplet of scalar fields ϕ_i which transform according to some representation, not necessarily irreducible, of G of dimension n. According to the rules we explained in the last section, the Lagrangian of the system is given by

$$\mathcal{L} = -\frac{1}{4}\mathrm{Tr}(G_{\mu\nu}G^{\mu\nu}) + (D_\mu\Phi)^\dagger D^\mu\Phi - V(\Phi)\ . \tag{54}$$

In component notation, the covariant derivative is, as usual, $D_\mu\phi_i = \partial_\mu\phi_i - \mathrm{i}g^{(a)}T^a_{ij}A^a_\mu\phi_j$ where we have allowed for the possibility of having arbitrary coupling constants $g^{(a)}$ for the various generators of G because we do not assume that G is simple or semi-simple. $V(\Phi)$ is a polynomial in the Φ invariant under G of degree equal to four. As before, we assume that we can choose the parameters in V such that the minimum is not at $\Phi = 0$ but rather at $\Phi = v$ where v is a constant vector in the representation space of Φ. v is not unique. The m generators of G can be separated into two classes: h generators which annihilate v and form the Lie algebra of the unbroken subgroup H; and $m - h$ generators, shown in the representation of Φ by matrices T^a, such that $T^a v \neq 0$ and all vectors $T^a v$ are independent and can be chosen to be orthogonal. Any vector in the orbit of v, i.e. of the form $\mathrm{e}^{\mathrm{i}w^a T^a}v$, is an equivalent minimum of the potential. As before, we should translate the scalar fields Φ by $\Phi \to \Phi + v$. It is convenient to decompose Φ into components along the orbit of v and orthogonal to it, the analogue of the χ and ψ fields of the previous section. We can write

$$\Phi = \mathrm{i}\sum_{a=1}^{m-h}\frac{\chi^a T^a v}{|T^a v|} + \sum_{b=1}^{n-m+h}\psi^b u^b + v\ , \tag{55}$$

where the vectors u^b form an orthonormal basis in the space orthogonal to all $T^a v$. The corresponding generators span the coset space G/H. As before, we shall show that the fields χ^a will be absorbed by the Brout–Englert–Higgs mechanism and the fields ψ^b will remain physical. Note that the set of vectors u^b contains at least one element since, for all a, we have

$$v \times T^a v = 0 \tag{56}$$

because the generators in a real unitary representation are anti-symmetric. This shows that the dimension n of the representation of Φ must be larger than $m - h$ and, therefore, there will remain at least one physical scalar field which, in the quantum theory, will give a physical scalar particle[7].

Let us now bring in the Lagrangian from Eq. (54) the expression of Φ from Eq. (55). We obtain

$$\mathcal{L} = \frac{1}{2}\sum_{a=1}^{m-h}(\partial_\mu\chi^a)^2 + \frac{1}{2}\sum_{b=1}^{n-m+h}(\partial_\mu\psi^b)^2 - \frac{1}{4}\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})$$
$$+\frac{1}{2}\sum_{a=1}^{m-h}g^{(a)2}|T^a v|^2 A^a_\mu A^{\mu a} - \sum_{a=1}^{m-h}g^{(a)}T^a v\partial^\mu\chi^a A^a_\mu - V(\Phi) + \cdots\ , \tag{57}$$

where the dots stand for coupling terms between the scalars and the gauge fields. In writing Eq. (57) we took into account that $T^b v = 0$ for $b > m - h$ and that the vectors $T^a v$ are orthogonal.

[7]Obviously, the argument assumes the existence of scalar fields which induce the phenomenon of spontaneous symmetry breaking. We can construct models in which the role of the latter is played by some kind of fermion–anti-fermion bound states and they come under the name of models with a *dynamical symmetry breaking*. In such models the existence of a physical spin-zero state, the analogue of the σ-particle of the chiral symmetry breaking of quantum chromodynamics (QCD), is a dynamical question and in general hard to answer.

The analysis that gave us Goldstone's theorem shows that

$$\frac{\partial^2 V}{\partial \phi_k \partial \phi_l}|_{\Phi=v}(T^a v)_l = 0 \; , \tag{58}$$

which shows that the χ-fields would correspond to the Goldstone modes. As a result, the only mass terms which appear in V in Eq. (57) are of the form $\psi^k M^{kl} \psi^l$ and do not involve the χ-fields.

As far as the bilinear terms in the fields are concerned, the Lagrangian from Eq. (57) is the sum of terms of the form found in the Abelian case. All gauge bosons which do not correspond to H generators acquire a mass equal to $m_a = g^{(a)}|T^a v|$ and, through their mixing with the would-be Goldstone fields χ, develop a zero-helicity state. All other gauge bosons remain massless. The ψ represent the remaining physical Higgs fields.

5 Building the Standard Model: a five-step programme

In this section we shall construct the Standard Model of electro-weak interactions as a spontaneously broken gauge theory. We shall follow the hints given by experiment following a five-step programme.

- Step 1: Choose a gauge group G.
- Step 2: Choose the fields of the 'elementary' particles and assign them to representations of G. Include scalar fields to allow for the Brout–Englert–Higgs mechanism.
- Step 3: Write the most general renormalizable Lagrangian invariant under G. At this stage, gauge invariance is still exact and all gauge vector bosons are massless.
- Step 4: Choose the parameters of the scalar potential so that spontaneous symmetry breaking occurs.
- Step 5: Translate the scalars and rewrite the Lagrangian in terms of the translated fields. Choose a suitable gauge and quantize the theory.

Note that gauge theories provide only the general framework, not a detailed model. The latter will depend on the particular choices made in Steps 1 and 2.

5.1 The lepton world

We start with the leptons and, in order to simplify the presentation, we shall assume that neutrinos are massless. We follow the five steps.

Step 1: Looking at the table of elementary particles we see that, for the combined electromagnetic and weak interactions, we have four gauge bosons, namely $W^{\pm}$, Z^0 and the photon. As we explained earlier, each one of them corresponds to a generator of the group G, more precisely its Lie algebra. The only non-trivial algebra with four generators is that of U(2) $\approx$ SU(2) $\times$ U(1).

Following the notation which was inspired by the hadronic physics, we call T_i, $i = 1, 2, 3$, the three generators of SU(2) and Y that of U(1). Then, the electric charge operator Q will be a linear combination of T_3 and Y. By convention, we write

$$Q = T_3 + \frac{1}{2} Y \; . \tag{59}$$

The coefficient in front of Y is arbitrary and only fixes the normalization of the U(1) generator relatively to those of SU(2)[8]. This ends our discussion of the first step.

[8]The normalization of the generators for non-Abelian groups is fixed by their commutation relations. That of the Abelian generator is arbitrary. The relation of Eq. (59) is one choice which has only a historical value. It is not the most natural one from the group theory point of view, as you will see in the discussion concerning Grand-Unified theories.

Step 2: The number and the interaction properties of the gauge bosons are fixed by the gauge group. This is no longer the case with the fields describing the other particles. In principle, we can choose any number and assign them to any representation. It follows that the choice here will be dictated by the phenomenology.

Leptons have always been considered as elementary particles. We have six leptons but, as we noted already, a striking feature of the data is the phenomenon of family repetition. We do not understand why nature chooses to repeat itself three times, but the simplest way to incorporate this observation into the model is to use the same representations three times, one for each family. This leaves SU(2) doublets and/or singlets as the only possible choices. A further experimental input we shall use is the fact that the charged W couple only to the left-handed components of the lepton fields, in contrast to the photon which couples with equal strength to both right and left. These considerations lead us to assign the left-handed components of the lepton fields to doublets of SU(2):

$$\Psi^{\mathrm{i}}_{\mathrm{L}}(x) = \frac{1}{2}(1+\gamma_5)\left(\begin{array}{c}\nu_i(x)\\ \ell^-_i(x)\end{array}\right)\ ;\quad i=1,2,3\ , \tag{60}$$

where we have used the same symbol for the particle and the associated Dirac field.

The right-handed components are assigned to singlets of SU(2):

$$\nu_{i\mathrm{R}}(x) = \frac{1}{2}(1-\gamma_5)\nu_i(x)\ \ (?)\ ;\quad \ell^-_{i\mathrm{R}}(x) = \frac{1}{2}(1-\gamma_5)\ell^-_i(x)\ . \tag{61}$$

The question mark next to the right-handed neutrinos means that the presence of these fields is not confirmed by the data. We shall drop them in this lecture, but we may come back to this point later. We shall also simplify the notation and put $\ell^-_{i\mathrm{R}}(x) = R_i(x)$. The resulting transformation properties under local SU(2) transformations are

$$\Psi^i_{\mathrm{L}}(x) \to \mathrm{e}^{\mathrm{i}\vec{\tau}\vec{\theta}(x)}\Psi^i_{\mathrm{L}}(x)\ ;\quad R_i(x)\to R_i(x)\ , \tag{62}$$

with $\vec{\tau}$ the three Pauli matrices. This assignment and the Y normalization given by Eq. (59), also fix the U(1) charge and, therefore, the transformation properties of the lepton fields. For all i we find

$$Y(\Psi^i_{\mathrm{L}}) = -1\ ;\quad Y(R_i) = -2\ . \tag{63}$$

If a right-handed neutrino exists, it has $Y(\nu_{i\mathrm{R}}) = 0$, which shows that it is not coupled to any gauge boson.

We are left with the choice of the Higgs scalar fields and we shall choose the solution with the minimal number of fields. We must give masses to three vector gauge bosons and keep the fourth one massless. The latter will be identified with the photon. We recall that, for every vector boson acquiring mass, a scalar with the same quantum numbers decouples. At the end we shall remain with at least one physical, neutral, scalar field. It follows that the minimal number to start with is four, two charged and two neutral. We choose to put them, under SU(2), into a complex doublet:

$$\Phi = \left(\begin{array}{c}\phi^+\\ \phi^0\end{array}\right)\ ;\quad \Phi(x)\to \mathrm{e}^{\mathrm{i}\vec{\tau}\vec{\theta}(x)}\Phi(x)\ , \tag{64}$$

with the conjugate fields ϕ^- and ϕ^{0*} forming $\Phi^\dagger$. The U(1) charge of Φ is $Y(\Phi)=1$.

This ends our choices for the second step. At this point the model is complete. All further steps are purely technical and uniquely defined.

Step 3: What follows is straightforward algebra. We write the most general, renormalizable, Lagrangian, involving the fields of Eqs. (60), (61) and (64) invariant under gauge transformations of

SU(2) $\times$ U(1). We shall also assume the separate conservation of the three lepton numbers, leaving the discussion on the neutrino mixing to a specialized lecture. The requirement of renormalizability implies that all terms in the Lagrangian are monomials in the fields and their derivatives, and their canonical dimension is less than or equal to four. The result is

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}\vec{W}_{\mu\nu} \times \vec{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + |D_\mu\Phi|^2 - V(\Phi) \\ &+ \sum_{i=1}^{3}\left[\bar{\Psi}^i_{\mathrm{L}} \mathrm{i}\not{D}\Psi^i_{\mathrm{L}} + \bar{R}_i \mathrm{i}\not{D}R_i - G_i(\bar{\Psi}^i_{\mathrm{L}} R_i \Phi + \mathrm{h.c.})\right] . \end{aligned} \tag{65}$$

If we call $\vec{W}$ and B the gauge fields associated with SU(2) and U(1) respectively, the corresponding field strengths $\vec{W}_{\mu\nu}$ and $B_{\mu\nu}$ appearing in Eq. (65) are given by Eqs. (24) and (15).

Similarly, the covariant derivatives in Eq. (65) are determined by the assumed transformation properties of the fields, as shown in Eq. (21):

$$\begin{gathered} D_\mu\Psi^i_{\mathrm{L}} = \left(\partial_\mu - \mathrm{i}g\tfrac{\vec{\tau}}{2} \times \vec{W}_\mu + \mathrm{i}\tfrac{g'}{2}B_\mu\right)\Psi^i_{\mathrm{L}} \ ; \ \ D_\mu R_i = \left(\partial_\mu + \mathrm{i}g'B_\mu\right)R_i, \\ D_\mu\Phi = \left(\partial_\mu - \mathrm{i}g\tfrac{\vec{\tau}}{2} \times \vec{W}_\mu - \mathrm{i}\tfrac{g'}{2}B_\mu\right)\Phi \ . \end{gathered} \tag{66}$$

The two coupling constants g and g' correspond to the groups SU(2) and U(1), respectively. The most general potential $V(\Phi)$ compatible with the transformation properties of the field Φ is

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \ . \tag{67}$$

The last term in Eq. (65) is a Yukawa coupling term between the scalar Φ and the fermions. In the absence of right-handed neutrinos, this is the most general term which is invariant under SU(2) $\times$ U(1). As usual, h.c. stands for 'hermitian conjugate'. G_i are three arbitrary coupling constants. If right-handed neutrinos exist there is a second Yukawa term with R_i replaced by ν_{iR} and Φ by the corresponding doublet proportional to $\tau_2\Phi^*$, where * means 'complex conjugation'. We see that the Standard Model can perfectly well accommodate a right-handed neutrino, but it couples only to the Higgs field.

A final remark: as expected, the gauge bosons $\vec{\mathrm{W}}_\mu$ and B_μ appear to be massless. The same is true for all fermions. This is not surprising because the assumed different transformation properties of the right- and left-handed components forbid the appearance of a Dirac mass term in the Lagrangian. On the other hand, the Standard Model quantum numbers also forbid the appearance of a Majorana mass term for the neutrinos. In fact, the only dimensionful parameter in (65) is μ^2, the parameter in the Higgs potential in Eq. (67). Therefore, the mass of every particle in the model is expected to be proportional to $|\mu|$.

Step 4: The next step of our programme consists of choosing the parameter μ^2 negative to trigger the phenomenon of spontaneous symmetry breaking and the Brout–Englert–Higgs mechanism. The minimum potential occurs at a point $v^2 = -\mu^2/\lambda$. As we have explained earlier, we can choose the direction of the breaking to be along the real part of ϕ^0.

Step 5: Translating the scalar field by a real constant,

$$\Phi \to \Phi + \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v^2 = -\frac{\mu^2}{\lambda} \ , \tag{68}$$

transforms the Lagrangian and generates new terms, as it was explained in the previous section. Let us look at some of them.

(i) *Fermion mass terms.* Replacing ϕ^0 by v in the Yukawa term in Eq. (65) creates a mass term for the charged leptons, leaving the neutrinos massless:

$$m_{\mathrm{e}} = \frac{1}{\sqrt{2}} G_{\mathrm{e}} v, \quad m_\mu = \frac{1}{\sqrt{2}} G_\mu v, \quad m_\tau = \frac{1}{\sqrt{2}} G_\tau v \ . \tag{69}$$

Since we have three arbitrary constants G_i, we can fit the three observed lepton masses. If we introduce right-handed neutrinos we can also apply whichever Dirac neutrino masses we wish.

(ii) *Gauge-boson mass terms.* They come from the $|D_\mu \Phi|^2$ term in the Lagrangian. A straight substitution produces the following quadratic terms among the gauge boson fields:

$$\frac{1}{8} v^2 [g^2 (W_\mu^1 W^{1\mu} + W_\mu^2 W^{2\mu}) + (g' B_\mu - g W_\mu^3)^2] \ . \tag{70}$$

Defining the charged vector bosons as

$$W_\mu^\pm = \frac{W_\mu^1 \mp \mathrm{i} W_\mu^2}{\sqrt{2}} \ , \tag{71}$$

we obtain their masses,

$$m_W = \frac{vg}{2} \ . \tag{72}$$

The neutral gauge bosons B_μ and W_μ^3 have a 2×2 non-diagonal mass matrix. After diagonalization, we define the mass eigenstates as

$$\begin{aligned} Z_\mu &= \cos\theta_{\mathrm{W}} B_\mu - \sin\theta_W W_\mu^3 \\ A_\mu &= \cos\theta_{\mathrm{W}} B_\mu + \sin\theta_W W_\mu^3 \ , \end{aligned} \tag{73}$$

with $\tan\theta_{\mathrm{W}} = g'/g$. They correspond to the mass eigenvalues

$$\begin{aligned} m_{\mathrm{Z}} &= \frac{v(g^2 + g'^2)^{1/2}}{2} = \frac{m_{\mathrm{W}}}{\cos\theta_{\mathrm{W}}} \\ m_A &= 0 \ . \end{aligned} \tag{74}$$

As expected, one of the neutral gauge bosons is massless and will be identified with the photon. The Brout–Englert–Higgs mechanism breaks the original symmetry according to SU(2) $\times$ U(1) $\rightarrow$ U(1)$_{em}$ and θ_{W} is the angle between the original U(1) and the one left unbroken. It is the parameter first introduced by S.L. Glashow, although it is often referred to as the 'Weinberg angle'.

(iii) *Physical Higgs mass.* Three out of the four real fields of the Φ doublet will be absorbed in order to allow for the three gauge bosons $\mathrm{W}^\pm$ and Z^0 to acquire a mass. The fourth one, which corresponds to $(|\phi^0 \phi^{0\dagger}|)^{1/2}$, remains physical. Its mass is given by the coefficient of the quadratic part of $V(\Phi)$ after the translation of Eq. (68) and is equal to

$$m_{\mathrm{h}} = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2} \ . \tag{75}$$

In addition, we produce various coupling terms which we shall present, together with the hadronic ones, in the next section.

5.2 Extension to hadrons

Introducing the hadrons into the model presents some novel features largely because the individual quark quantum numbers are not separately conserved. With regard to the second step, there is currently a consensus regarding the choice of the 'elementary' constituents of matter: besides the six leptons, there are six quarks. They are fractionally charged and come each in three 'colours'. The observed lepton–hadron universality property tells us to also use doublets and singlets for the quarks. The first novel feature we mentioned above is that all quarks appear to have non-vanishing Dirac masses, so we must introduce both right-handed singlets for each family. A naïve assignment would be to write the analogue of Eqs. (60) and (61) as

$$Q_{\rm L}^i(x)=\frac{1}{2}(1+\gamma_5)\begin{pmatrix}U^i(x)\\ D^i(x)\end{pmatrix}\;;\quad U_{\rm R}^i(x)\;;\quad D_{\rm R}^i(x)\;, \tag{76}$$

with the index i running over the three families as U^i =u,c,t and D^i =d,s,b for $i=1,2,3$, respectively[9]. This assignment determines the SU(2) transformation properties of the quark fields. It also fixes their Y charges and, hence their U(1) properties. Using Eq. (59), we find

$$Y(Q_{\rm L}^i)=\frac{1}{3}\;;\quad Y(U_{\rm R}^i)=\frac{4}{3}\;;\quad Y(D_{\rm R}^i)=-\frac{2}{3}\;. \tag{77}$$

The presence of the two right-handed singlets has an important consequence. Even if we had only one family, we would have two distinct Yukawa terms between the quarks and the scalar field of the form

$$\mathcal{L}_{\rm Yuk}=G_{\rm d}(\bar{Q}_{\rm L}D_{\rm R}\Phi+\text{h.c.})+G_{\rm u}(\bar{Q}_{\rm L}U_{\rm R}\tilde{\Phi}+\text{h.c.})\;. \tag{78}$$

$\tilde{\Phi}$ is the doublet proportional to $\tau_2\Phi^*$. It has the same transformation properties under SU(2) as Φ, but the opposite Y charge.

If there were only one family, this would have been the end of the story. The hadron Lagrangian $\mathcal{L}_{\rm h}^{(1)}$ is the same as Eq. (65) with quark fields replacing leptons and the extra term of Eq. (78). The complication we alluded to before comes with the addition of more families. In this case the total Lagrangian is not just the sum over the family index. The physical reason is the non-conservation of the individual quark quantum numbers we mentioned previously. In writing Eq. (76), we implicitly assumed a particular pairing of the quarks in each family; u with d, c with s and t with b. In general, we could choose any basis in family space and, since we have two Yukawa terms, we will not be able to diagonalize both of them simultaneously. It follows that the most general Lagrangian will contain a matrix with non-diagonal terms which mix the families. By convention, we attribute it to a different choice of basis in the d–s–b space. It follows that the correct generalization of the Yukawa Lagrangian of Eq. (78) to many families is given by

$$\mathcal{L}_{\rm Yuk}=\sum_{i,j}\left[(\bar{Q}_{\rm L}^iG_{\rm d}^{ij}D_{\rm R}^j\Phi+\text{h.c.})\right]+\sum_i\left[G_{\rm u}^i(\bar{Q}_{\rm L}^iU_{\rm R}^i\tilde{\Phi}+\text{h.c.})\right]\;, \tag{79}$$

where the Yukawa coupling constant $G_{\rm d}$ has become a matrix in family space. After translation of the scalar field, we shall produce masses for the up quarks given by $m_{\rm u}=G_{\rm u}^1v$, $m_{\rm c}=G_{\rm u}^2v$ and $m_{\rm t}=G_{\rm u}^3v$, as well as a 3×3 mass matrix for the down quarks given by $G_{\rm d}^{ij}v$. As usual, we want to work in a field space where the masses are diagonal, so we change our initial d–s–b basis to bring $G_{\rm d}^{ij}$ into a diagonal form. This can be done through a 3×3 unitary matrix $\tilde{D}^i=U^{ij}D^j$ such that $U^\dagger G_{\rm d}U=$ diag$(m_{\rm d},m_{\rm s},m_{\rm b})$. In the simplest example of only two families, it is easy to show that the most general such matrix, after using all freedom for field redefinitions and phase choices, is a real rotation:

[9] An additional index a, also running through 1, 2 and 3 and denoting the colour, is understood.

$$C = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} , \tag{80}$$

with θ being our familiar Cabibbo angle. For three families, an easy counting shows that the matrix has three angles, the three Euler angles, and an arbitrary phase. It is traditionally written in the form

$$KM = \begin{pmatrix} c_1 & s_1c_3 & s_1s_3 \\ -s_1c_3 & c_1c_2c_3 - s_2s_3\mathrm{e}^{\mathrm{i}\delta} & c_1c_2s_3 + s_2c_3\mathrm{e}^{\mathrm{i}\delta} \\ -s_1s_2 & c_1s_2c_3 + c_2s_3\mathrm{e}^{\mathrm{i}\delta} & c_1s_2s_3 - c_2c_3\mathrm{e}^{\mathrm{i}\delta} \end{pmatrix} , \tag{81}$$

with the notation $c_k = \cos\theta_k$ and $s_k = \sin\theta_k$, $k = 1, 2, 3$. The novel feature is the possibility of introducing the phase δ. This means that a six-quark model has a natural source of CP or T violation, whereas a four-quark model does not.

The total Lagrangian density, before the translation of the field Φ, is now

$$\begin{aligned} \mathcal{L} \;=\; & -\frac{1}{4}\vec{W}_{\mu\nu} \times \vec{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + |D_\mu\Phi|^2 - V(\Phi) \\ & + \sum_{i=1}^{3} \left[\bar{\Psi}^i_{\mathrm{L}}\mathrm{i}\not{D}\Psi^i_{\mathrm{L}} + \bar{R}_i\mathrm{i}\not{D}R_i - G_i(\bar{\Psi}^i_{\mathrm{L}}R_i\Phi + \mathrm{h.c.})\right. \\ & + \left.\bar{Q}^i_{\mathrm{L}}\mathrm{i}\not{D}Q^i_{\mathrm{L}} + \bar{U}^i_{\mathrm{R}}\mathrm{i}\not{D}U^i_{\mathrm{R}} + \bar{D}^i_{\mathrm{R}}\mathrm{i}\not{D}D^i_{\mathrm{R}} + G^i_{\mathrm{u}}(\bar{Q}^i_{\mathrm{L}}U^i_{\mathrm{R}}\tilde{\Phi} + \mathrm{h.c.})\right] \\ & + \sum_{i,j=1}^{3} \left[(\bar{Q}^i_{\mathrm{L}}G^{ij}_{\mathrm{d}}D^j_{\mathrm{R}}\Phi + \mathrm{h.c.})\right] . \end{aligned} \tag{82}$$

The covariant derivatives on the quark fields are given by

$$\begin{aligned} D_\mu Q^i_{\mathrm{L}} &= \left(\partial_\mu - \mathrm{i}g\frac{\vec{\tau}}{2} \times \vec{W}_\mu - \mathrm{i}\frac{g'}{6}B_\mu\right) Q^i_{\mathrm{L}} \\ D_\mu U^i_{\mathrm{R}} &= \left(\partial_\mu - \mathrm{i}\frac{2g'}{3}B_\mu\right) U^i_{\mathrm{R}} \\ D_\mu D^i_{\mathrm{R}} &= \left(\partial_\mu + \mathrm{i}\frac{g'}{3}B_\mu\right) D^i_{\mathrm{R}} . \end{aligned} \tag{83}$$

The classical Lagrangian in Eq. (82) contains 17 arbitrary real parameters. They are:

- the two gauge coupling constants g and g';
- the two parameters of the scalar potential λ and μ^2;
- three Yukawa coupling constants for the three lepton families, $G_{e,\mu,\tau}$;
- six Yukawa coupling constants for the three quark families, $G^{u,c,t}_{\mathrm{u}}$; and $G^{d,s,b}_{\mathrm{d}}$.
- four parameters of the KM matrix, the three angles and the phase δ.

A final remark: 15 out of these 17 parameters are directly connected with the Higgs sector.

Translating the scalar field by Eq. (68) and diagonalizing the resulting down-quark mass matrix produces the mass terms for fermions and bosons as well as several coupling terms. We shall write here the ones which involve the physical fields[10].

[10] We know from QED that, in order to determine the Feynman rules of a gauge theory, one must first decide on a choice of gauge. For Yang–Mills theories, this step introduces new fields called *Faddeev–Popov ghosts*. This point is explained in every standard text book on quantum field theory, but we have not discussed it in these lectures.

(i) *The gauge boson–fermion couplings.* They are the ones which generate the known weak and electromagnetic interactions. A_μ is coupled to the charged fermions through the usual electromagnetic current:

$$\frac{gg'}{(g^2+g'^2)^{1/2}}\left[\bar{e}\gamma^\mu e+\sum_{a=1}^{3}\left(\frac{2}{3}\bar{u}^a\gamma^\mu u^a-\frac{1}{3}\bar{d}^a\gamma^\mu d^a\right)\ +\cdots\right]A_\mu\ , \tag{84}$$

where the dots stand for the contribution of the other two families e $\to \mu, \tau$, u $\to$ c,t and d $\to$ s,b and the summation over a extends over the three colours. Equation (84) shows that the electric charge e is given in terms of g and g' by

$$e=\frac{gg'}{(g^2+g'^2)^{1/2}}=g\sin\theta_{\rm W}=g'\cos\theta_{\rm W}\ . \tag{85}$$

Similarly, the couplings of the charged W to the weak current are

$$\frac{g}{2\sqrt{2}}\left(\bar{\nu}_e\gamma^\mu(1+\gamma_5)e+\sum_{a=1}^{3}\bar{u}^a\gamma^\mu(1+\gamma_5)d^a_{KM}\ \ +\cdots\right)W^+_\mu\ \ +\text{h.c.}\ . \tag{86}$$

Combining all these relations, we can determine the experimental value of the parameter v, the vacuum expectation value of the Higgs field. We find $v\sim 246$ GeV.

As expected, only left-handed fermions participate. d_{KM} is the linear combination of d–s–b given by the KM matrix in Eq. (81). By diagonalizing the down-quark mass matrix, we introduced the off-diagonal terms into the hadron current. When considering processes, like nuclear β-decay or μ-decay, where the momentum transfer is very small compared to the W mass, the W propagator can be approximated by ${m_{\rm W}}^{-2}$ and the effective Fermi coupling constant is given by

$$\frac{G}{\sqrt{2}}=\frac{g^2}{8m_{\rm W}^2}=\frac{1}{2v^2}\ . \tag{87}$$

In contrast to the charged weak current shown in Eq. (86), the Z^0-fermion couplings involve both left- and right-handed fermions:

$$\begin{aligned}-\frac{e}{2}\frac{1}{\sin\theta_{\rm W}\cos\theta_{\rm W}}\ &[\bar{\nu}_{\rm L}\gamma^\mu\nu_{\rm L}+(\sin^2\theta_{\rm W}-\cos^2\theta_{\rm W})\bar{e}_{\rm L}\gamma^\mu e_{\rm L}\\ &\qquad +2\sin^2\theta_{\rm W}\bar{e}_{\rm R}\gamma^\mu e_{\rm R}\ +\ \cdots]\ Z_\mu\ ,\end{aligned} \tag{88}$$

$$\begin{aligned}\frac{e}{2}\sum_{a=1}^{3}\Bigg[&\left(\frac{1}{3}\tan\theta_{\rm W}-\cot\theta_{\rm W}\right)\bar{u}^a_{\rm L}\gamma^\mu u^a_{\rm L}+\left(\frac{1}{3}\tan\theta_{\rm W}+\cot\theta_{\rm W}\right)\bar{d}^a_{\rm L}\gamma^\mu d^a_{\rm L}\\ &+\frac{2}{3}\tan\theta_{\rm W}(2\bar{u}^a_{\rm R}\gamma^\mu u^a_{\rm R}-\bar{d}^a_{\rm R}\gamma^\mu d^a_{\rm R})\ +\ \cdots\Bigg]\ Z_\mu\ .\end{aligned} \tag{89}$$

Again, the summation is over the colour indices and the dots stand for the contribution of the other two families. In this formula we verify the property of the weak neutral current to be diagonal in the quark-flavour space. Another interesting property is that the axial part of the neutral current is proportional to $[\bar{u}\gamma_\mu\gamma_5 u-\bar{d}\gamma_\mu\gamma_5 d]$. This particular form of the coupling is important for phenomenological applications, such as the induced parity violating effects in atoms and nuclei.

(ii) *The gauge boson self-couplings.* One of the characteristic features of Yang–Mills theories is the particular form of the self-couplings among the gauge bosons. They come from the square of the non-Abelian curvature in the Lagrangian, which, in our case, is the term $-\frac{1}{4}\vec{W}_{\mu\nu} \times \vec{W}^{\mu\nu}$. Expressed in terms of the physical fields, this term gives

$$\begin{aligned}
&- \mathrm{i}g(\sin\theta_{\mathrm{W}} A^{\mu} - \cos\theta_{\mathrm{W}} Z^{\mu})(W^{\nu -} W^{+}_{\mu\nu} - W^{\nu +} W^{-}_{\mu\nu}) \\
&- \mathrm{i}g(\sin\theta_{\mathrm{W}} F^{\mu\nu} - \cos\theta_{\mathrm{W}} Z^{\mu\nu}) W^{-}_{\mu} W^{+}_{\nu} \\
&- g^2(\sin\theta_{\mathrm{W}} A^{\mu} - \cos\theta_{\mathrm{W}} Z^{\mu})^2 W^{+}_{\nu} W^{\nu -} \\
&+ g^2(\sin\theta_{\mathrm{W}} A^{\mu} - \cos\theta_{\mathrm{W}} Z^{\mu})(\sin\theta_{\mathrm{W}} A^{\nu} - \cos\theta_{\mathrm{W}} Z^{\nu}) W^{+}_{\mu} W^{-}_{\nu} \\
&- \frac{g^2}{2}(W^{+}_{\mu} W^{\mu -})^2 + \frac{g^2}{2}(W^{+}_{\mu} W^{-}_{\nu})^2 \,,
\end{aligned} \tag{90}$$

where we have used the following notation: $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $W^{\pm}_{\mu\nu} = \partial_\mu W^{\pm}_{\nu} - \partial_\nu W^{\pm}_{\mu}$ and $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$ with $g\sin\theta_{\mathrm{W}} = e$. Let us concentrate on the photon–$\mathrm{W}^+\mathrm{W}^-$ couplings. If we forget, for the moment, about the SU(2) gauge invariance, we can use different coupling constants for the two trilinear couplings in Eq. (90), say e for the first and $e\kappa$ for the second. For a charged, massive W, the magnetic moment μ and the quadrupole moment Q are given by

$$\mu = \frac{(1+\kappa)e}{2m_{\mathrm{W}}} \qquad Q = -\frac{e\kappa}{m_{\mathrm{W}}^2} \,. \tag{91}$$

Looking at Eq. (90), we see that $\kappa = 1$. Therefore, SU(2) gauge invariance gives very specific predictions concerning the electromagnetic parameters of the charged vector bosons. The gyromagnetic ratio equals two and the quadrupole moment equals $-em_{\mathrm{W}}^{-2}$.

(iii) *The scalar fermion couplings.* They are given by the Yukawa terms in Eq. (65). The same couplings generate the fermion masses through spontaneous symmetry breaking. It follows that the physical Higgs scalar couples to quarks and leptons with strength proportional to the fermion mass. Therefore, the prediction is that it will decay predominantly to the heaviest possible fermion compatible with phase space. This property provides a typical signature for its identification.

(iv) *The scalar gauge boson couplings.* They come from the covariant derivative term $|D_\mu\Phi|^2$ in the Lagrangian. If we call ϕ the field of the physical neutral Higgs, we find

$$\frac{1}{4}(v+\phi)^2\left[g^2 W^{+}_{\mu} W^{-\mu} + (g^2 + g'^2) Z_\mu Z^\mu\right] \,. \tag{92}$$

This gives a direct coupling ϕ–W^+–W^-, as well as ϕ–Z–Z, which has been very useful in the Higgs searches.

(v) *The scalar self-couplings.* They are proportional to $\lambda(v+\phi)^4$. Equations (75) and (87) show that $\lambda = Gm_{\mathrm{h}}^2/\sqrt{2}$, so, in the tree approximation, this coupling is related to the Higgs mass. It could provide a test of the Standard Model Higgs, but it will not be easy to measure. On the other hand, this relation shows that, were the physical Higgs very heavy, it would also have been strongly interacting, and this sector of the model would become non-perturbative.

The five-step programme is now complete for both leptons and quarks. The 17 parameters of the model have all been determined by experiment. Although the number of arbitrary parameters seems very large, we should not forget that they are all mass and coupling parameters, like the electron mass and the fine structure constant of QED. The reason we have more of them is that the Standard Model describes a much larger number of particles and interactions in a unified framework .

6 The Standard Model and experiment

Our confidence in this model is amply justified on the basis of its ability to accurately describe the bulk of our present-day data and, especially, of its enormous success in predicting new phenomena. Let us mention a few of them. We shall follow the historical order.

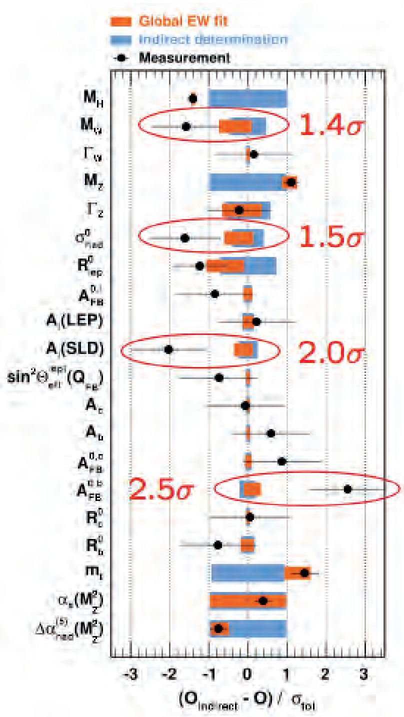

Fig. 5: A comparison between measured and computed values for various physical quantities

- The discovery of weak neutral currents by Gargamelle in 1972:

  $\nu_\mu + \mathrm{e}^- \to \nu_\mu + \mathrm{e}^- \; ; \quad \nu_\mu + \mathrm{N} \to \nu_\mu + \mathrm{X} \, .$

  Both their strength and their properties were predicted by the Standard Model.
- The discovery of charmed particles at SLAC in 1974. Their presence was essential to ensure the absence of strangeness changing neutral currents, for example $\mathrm{K}^0 \to \mu^+ + \mu^-$. Their characteristic property is to decay predominantly into strange particles.
- A necessary condition for the consistency of the Model is that $\sum_i Q_i = 0$ inside each family. When the τ lepton was discovered this implied a prediction for the existence of the b and t quarks with the right electric charges.
- The observed CP violation could be naturally incorporated into a model with three families. The b and t quarks were indeed discovered.
- The discovery of the W and Z bosons at CERN in 1983 with the masses predicted by the theory. The characteristic relation of the Standard Model with an isodoublet Brout–Englert–Higgs mechanism $m_{\mathrm{Z}} = m_{\mathrm{W}} / \cos\theta_{\mathrm{W}}$ has been checked with very high accuracy (including radiative corrections).
- The t-quark was *seen* at LEP through its effects in radiative corrections before its actual discovery at Fermilab.
- The vector boson self-couplings, γ–W^+–W^- and Z^0–W^+–W^- have been measured at LEP and confirm the Yang–Mills predictions given in Eq. (91).
- The recent discovery of a new boson which can be identified with the Higgs particle of the Standard Model is the last of this impressive series of successes.

All these discoveries should not make us forget that the Standard Model has been equally successful in fitting a large number of experimental results. You have all seen the global fit given in Fig. 5. The conclusion is obvious: *the Standard Model has been enormously successful.*

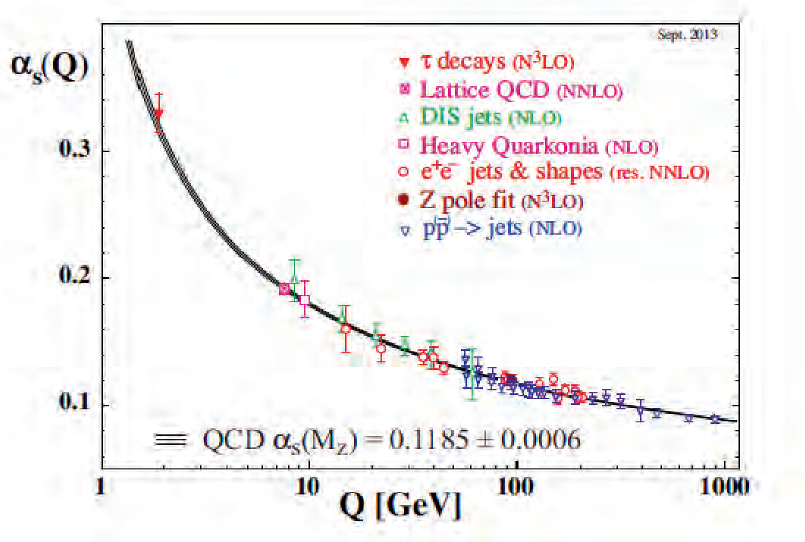

Fig. 6: The effective coupling constant for strong interactions as a function of the energy scale

Although in these lectures we did not discuss QCD, the gauge theory of strong interactions, the computations whose results are presented in Fig. 5, take into account the radiative corrections induced by virtual gluon exchanges. The fundamental property of QCD, the one which allows for perturbation theory calculations, is the property of asymptotic freedom, which is the particular dependence of the effective coupling strength on the energy scale. This is presented in Fig. 6 which shows the theoretical prediction based on QCD calculations, including the theoretical uncertainties. We see that the agreement with the experimentally measured values of the effective strong interaction coupling constant α_{s} is truly remarkable. Notice that this agreement extends to rather low values of Q of the order of 1–2 GeV, where α_{s} equals approximately $1/3$.

This brings us to our next point, namely that the success presented so far is in fact *a success of renormalized perturbation theory.* The extreme accuracy of the experimental measurements, mainly at LEP but also at FermiLab and elsewhere, allow a detailed comparison between theory and experiment to be made for the first time *including the purely weak interaction radiative corrections.*

In Fig. 7 we show the comparison between theory and experiment for two quantities, ϵ_1 and ϵ_3, defined in Eqs. (93) and (94), respectively:

$$\epsilon_1 = \frac{3G_F m_{\mathrm{t}}^2}{8\sqrt{2}\pi^2} - \frac{3G_F m_{\mathrm{W}}^2}{4\sqrt{2}\pi^2} \tan^2\theta_{\mathrm{W}} \ln\frac{m_{\mathrm{H}}}{m_{\mathrm{Z}}} + \cdots , \tag{93}$$

$$\epsilon_3 = \frac{G_F m_{\mathrm{W}}^2}{12\sqrt{2}\pi^2} \ln\frac{m_{\mathrm{H}}}{m_{\mathrm{Z}}} - \frac{G_F m_{\mathrm{W}}^2}{6\sqrt{2}\pi^2} \ln\frac{m_{\mathrm{t}}}{m_{\mathrm{Z}}} + \cdots . \tag{94}$$

They are defined with the following properties: (i) they include the strong and electromagnetic radiative corrections; and (ii) they vanish in the Born approximation for the weak interactions. So, they measure the weak interaction radiative corrections. The figure shows that, in order to obtain agreement with the data, one must include these corrections. Weak interactions are no longer a simple phenomenological model, but have become a precision theory.

The moral of the story is that the perturbation expansion of the Standard Model is reliable as long as all coupling constants remain small. The only coupling which does become large in some kinematical regions is α_{s}, which grows at small energy scales, as shown in Fig. 6. In this region, we know that a hadronization process occurs and perturbation theory breaks down. New techniques are necessary in

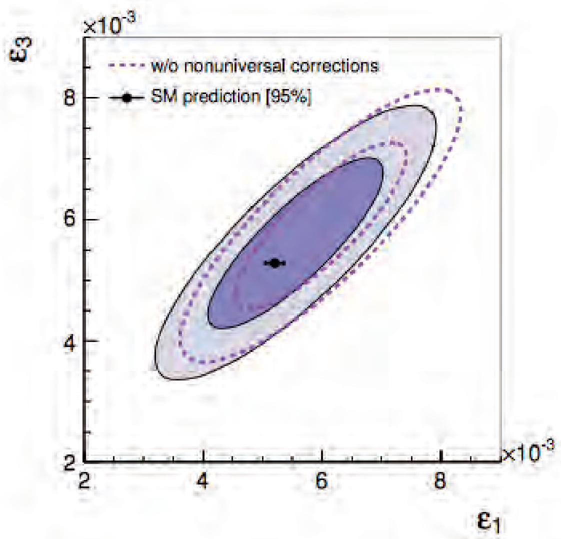

Fig. 7: Comparison between theory and experiment for two quantities sensitive to weak interaction radiative corrections.

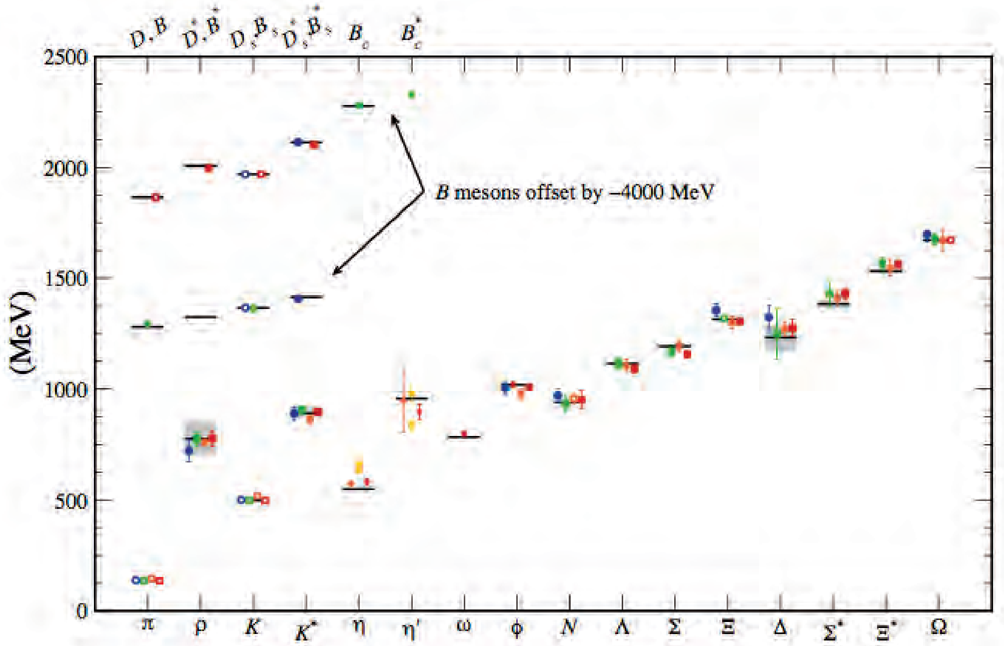

Fig. 8: The hadron spectrum obtained by numerical simulations of QCD on a space–time lattice

order to compare theoretical predictions with experimental data. In recent years, considerable effort has been devoted to this question with extensive numerical studies of QCD in the approximation in which the four-dimensional space–time has been replaced by a finite lattice. In Fig. 8 we show the computed spectrum of low-lying hadron states and the comparison with the data. The agreement makes us believe that we control the theory at both the weak- and strong-coupling regime. We should no longer talk about the Standard Model, but rather about the Standard Theory of the interactions among elementary particles. As a by-product of this analysis, we feel confident to say that at high energies perturbation theory is expected to be reliable *unless there are new strong interactions.*

This brings us to our last point that this very success shows that the Standard Model cannot be a complete theory. In other words there must be new physics beyond the Standard Model. The argument is simple and it is based on a straightforward application of perturbation theory with an additional assumption which we shall explain presently.

We assume that the Standard Model is correct up to a certain scale Λ. The precise value of Λ does not matter, provided it is larger than any energy scale reached so far[11].

A quantum field theory is defined through a functional integral over all classical field configurations, the Feynman path integral. By a Fourier transformation we can express it as an integral over the fields defined in momentum space. Following K. Wilson, let us split this integral in two parts: the high-energy part with modes above Λ and the low-energy part with the modes below Λ. Let us imagine that we perform the high-energy part. The result will be an effective theory expressed in terms of the low-energy modes of the fields. We do not know how to perform this integration explicitly, so we cannot write down the correct low-energy theory, but the most general form will be a series of operators made out of powers of the fields and their derivatives. Since integrating over the heavy modes does not break any of the symmetries of the initial Lagrangian, only operators allowed by the symmetries will appear. Wilson remarked that, when Λ is large compared to the mass parameters of the theory, we can determine the leading contributions by simple dimensional analysis[12]. We distinguish three kinds of operators, according to their canonical dimension.

– Those with dimension larger than four. Dimensional analysis shows that they will come with a coefficient proportional to inverse powers of Λ, so, by choosing a scale large enough, we can make their contribution arbitrarily small. We shall call them *irrelevant operators*.
– Those with dimension equal to four. They are the ones which appeared already in the original Lagrangian. Their coefficient will be independent of Λ, up to logarithmic corrections which we ignore. We shall call them *marginal operators*.
– Finally, we have the operators with dimension smaller than four. In the Standard Model there is only one such operator, the square of the scalar field Φ^2 which has dimension equal to two[13]. This operator will appear with a coefficient proportional to Λ^2, which means that its contribution will grow quadratically with Λ. We shall call it the *relevant operator*. It will give an effective mass to the scalar field proportional to the square of whichever scale we can think of. This problem was first identified in the framework of Grand Unified Theories and is known since as *the hierarchy problem*. Let me emphasize here that this does not mean that the mass of the scalar particle will be necessarily equal to Λ. The Standard Model is a renormalizable theory and the mass is fixed by a renormalization condition to its physical value. It only means that this condition should be adjusted to arbitrary precision order by order in perturbation theory. It is this extreme sensitivity to high scales, known as *the fine tuning problem*, which is considered unacceptable for a fundamental theory.

Let us summarize: the great success of the Standard Model tells us that renormalized perturbation theory is reliable in the absence of strong interactions. The same perturbation theory shows the need of a fine tuning for the mass of the scalar particle. If we do not accept the latter, we have the following two options.

[11]The scale Λ should not be confused with a cut-off that is often introduced when computing Feynman diagrams. This cut-off disappears after renormalization is performed. Here Λ is a physical scale which indicates how far the theory can be trusted.

[12]There are some additional technical assumptions concerning the dimensions of the fields, but they are satisfied in perturbation theory.

[13]There exists also the unit operator with dimension equal to zero which induces an effective cosmological constant. Its effects are not observable in a theory which ignores the gravitational interactions, so we shall not discuss it here. One could think of the square of a fermion operator $\bar{\Psi}\Psi$, whose dimension is equal to three, but it is not allowed by the chiral symmetry of the model.

– Perturbation theory breaks down at some scale Λ. We can imagine several reasons for a such a breakdown to occur. The simplest is the appearance of new strong interactions. The so-called *technicolor* models, in which the role of the Higgs field is played by a bound state of new strongly coupled fermions, were in this class. More exotic possibilities include the appearance of new, compact space dimensions with compactification length $\sim \Lambda^{-1}$.
– Perturbation theory is still valid but the numerical coefficient of the Λ^2 term which multiplies the Φ^2 operator vanishes to all orders of perturbation theory. For this to happen we must modify the Standard Model introducing appropriate new particles. Supersymmetry is the only systematic way we know to achieve this goal.

7 Conclusions

In these lectures we saw the fundamental role of geometry in the dynamics of the forces among the elementary particles. It was the understanding of this role which revolutionized our way of thinking and led to the construction of the Standard Model. It incorporates the ideas of gauge theories, as well as those of spontaneous symmetry breaking. Its agreement with experiment is spectacular. It fits all data known today. However, unless one is willing to accept a fine tuning with arbitrary precision, one should conclude that new physics will appear beyond a scale Λ. The precise value of Λ cannot be computed, but the amount of fine tuning grows quadratically with it, so it cannot be too large. Hopefully, it will be within reach of the LHC.

Appendix A: The principles of renormalization

In this appendix I want to recall and summarize the basic principles of perturbative renormalization theory. Since renormalization has a well-deseved reputation of complexity, this will be done by omitting all technical details. My purpose is to dissipate a widely spread belief according to which renormalization is a mathematically murky procedure: adding and subtracting infinities. On the contrary, I want to explain that it offers the only known mathematically consistent way to define the perturbation expansion of a quantum field theory.

A.1 The need for renormalization

Everyone who has attempted to compute a one-loop Feynman diagram knows that divergent expressions are often encountered. For example, in the ϕ^4 theory we find the diagram of Fig. A.1 involving the integral

$$I = \int \frac{\mathrm{d}^4 k}{(k^2 - m^2 + \mathrm{i}\epsilon)[(k-p)^2 - m^2 + \mathrm{i}\epsilon]} \, , \tag{A1}$$

which diverges logarithmically at large k. Similar divergences can be found in any theory, such as QED, Yang–Mills, etc. They have no place in a well-defined mathematical theory. So, if we find them, it means that we have made a mathematical mistake somewhere. Where is it? Let us first notice that the divergence in Eq. (A1) occurs at large values of the internal momentum, which, by Fourier transform, implies short distances. Did we make a mistake at short distances? Yes we did! We wrote the Lagrangian density as

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi(x)\right)\left(\partial^\mu \phi(x)\right) - \frac{1}{2} m^2 (\phi(x))^2 - \frac{\lambda}{4!}(\phi(x))^4 \, . \tag{A2}$$

On the other hand, the canonical commutation relations for a scalar quantum field are given by

$$\left[\phi(\vec{x}, t), \dot{\phi}(\vec{y}, t)\right] = \mathrm{i}\hbar \delta^3(\vec{x} - \vec{y}) \, . \tag{A3}$$

We know that the Dirac δ-function is not really a 'function' but a special form of what we call 'a distribution'. Many properties of well-behaved functions do not apply to it. In particular, the multiplication is not always a well-defined operation. $(\delta(x))^2$ is meaningless. The presence of the δ-function on the right-hand side of Eq. (A3) implies that the field $\phi(x)$ is also a distribution[14], so the product ϕ^2 is ill defined. Yet, it is precisely expressions of this kind that we wrote in every single term of our Lagrangian Eq. (A2). Since our initial Lagrangian is not well defined, it is not surprising that our calculations yield divergent results.

Now that we have identified the origin of the problem, we can figure out ways to solve it. A conceptually simple one would be to replace the field products in Eq. (A2) by splitting the points:

$$\phi(x)\phi(x) \to \lim_{a \to 0} \phi\left(x + \frac{a}{2}\right) \phi\left(x - \frac{a}{2}\right) \, . \tag{A4}$$

This expression is perfectly well defined for all values of the parameter a, except $a = 0$. In terms of distributions this means that the product is defined up to an arbitrary distribution $\mathcal{F}(a)$ which has support (i.e. it is non-zero), only at $a = 0$. Such a distribution is a superposition of the δ-function and its derivatives,

$$\mathcal{F}(a) = \sum_i C_i \delta^{(i)}(a) \tag{A5}$$

[14] The precise term is 'operator valued distribution'.

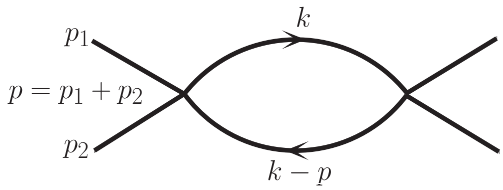

Fig. A.1: An one-loop divergent diagram in the ϕ^4 theory

with the C_i arbitrary real constants. The moral of the story is that the quantization rules for a local-field theory imply that every term in the Lagrangian contains a set of arbitrary constants which must be determined by experiment. Renormalization is the mathematical procedure which allows us to do it. A final remark: how many parameters are needed in order to define a given field theory? The answer involves the distinction between *renormalizable* and *non-renormalizable* theories. For the first, a finite number suffices. For the second, we need an infinite number, which means that non-renormalizable theories have no predictive power.

A.2 The theory of renormalization

In this section, I want to give some more information concerning the renormalization prescription. The process we outlined above was formulated in x-space. It is intuitively easier to understand, but not very convenient for practical calculations, which are usually performed in momentum space. The connection is by Fourier transform. The derivatives of the δ-function in Eq. (A5) become polynomials in the external momenta.

The renormalization programme follows three steps:

– *the power counting* which determines how many constants C we shall need for a given field theory;
– *the regularization* which is a prescription to make every Feynman diagram finite with the price of introducing a new parameter in the theory, the analogue of the point-splitting parameter a we used in Eq. (A4);
– *the renormalization* which is the mathematical procedure to eliminate the regularization parameter and determine the values of the necessary constants C.

A.2.1 The power counting

As the term indicates, it is the counting which determines whether a given diagram is divergent or not. We shall need to introduce some terminology. First, we have the obvious notions of *disconnected* and *connected* diagrams. A further specification is the *one-particle irreducible* (1PI) diagrams. A diagram is 1PI if it cannot be separated into two disconnected pieces by cutting a single internal line. A general connected diagram is constructed by joining together 1PI pieces, see Fig. A.2. It is obvious that a connected diagram is divergent if, and only if, one or more of its 1PI pieces is divergent, because the momenta of the internal connecting lines are fixed by energy-momentum conservation in terms of the external momenta and bring no new integrations.

This brings us to the power-counting argument. A single loop integral will be ultravioletly divergent if and only if the numerator is of equal or higher degree in the loop momentum than the denominator. For multiloop diagrams this may not be the case, since the divergence may be entirely due to a particular sub-diagram. However, in the spirit of perturbation theory, the divergent sub-diagram must be treated first. We thus arrive at the notion of *superficial degree of divergence* d of a given 1PI diagram, defined as the difference between the degree of integration momenta of the numerator minus that of the denominator. The diagram will be called *primitively divergent* if $d \geq 0$. Let us compute, as an example, d for

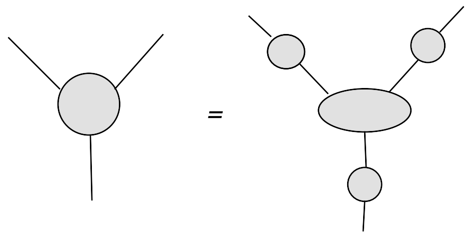

Fig. A.2: The 1PI decomposition of the three point function

the diagrams of the scalar field theory described in Eq. (A2), in the generalization in which we replace the interaction term ϕ^4 by ϕ^m with m integer, $m \geq 3$. Let us consider an 1PI diagram of nth order in perturbation with I internal and E external lines. Every internal line brings four powers of k to the numerator through the d^4k factor and two powers in the denominator through the propagator. Every vertex brings a δ^4-function of the energy-momentum conservation. All but one of them can be used to eliminate one integration each, the last reflecting the overall conservation which involves only external momenta. Therefore, we obtain

$$d = 2I - 4n + 4\ . \tag{A6}$$

This expression can be made more transparent by expressing I in terms of E and m. A simple counting gives $2I + E = mn$ and Eq. (A6) becomes

$$d = (m-4)n - E + 4\ . \tag{A7}$$

This is the main result. Although it is shown here as a plausibility argument, it is in fact a rigorous result. We see that $m = 4$ is a critical value and we can distinguish three cases.

1. $m = 3$, $d = 4 - n - E$. d is a decreasing function of n, the order of perturbation theory. Only a limited number of diagrams are primitively divergent. Above a certain order they are all convergent. For reasons that will be clear soon, we shall call such theories *super-renormalizable*.
2. $m = 4$, $d = 4 - E$. d is independent of the order of perturbation theory. If a Green function is divergent at some order, it will be divergent at all orders. For the ϕ^4 theory we see that the primitively divergent diagrams are those with $E = 2$, which have $d = 2$ and are quadratically divergent and those with $E = 4$ which have $d = 0$ and are logarithmically divergent. (Notice that, for this theory, all Green functions with odd E vanish identically because of the symmetry $\phi \to -\phi$). We shall call such theories *renormalizable*.
3. $m > 4$, d is an increasing function of n. Every Green function, irrespective of the number of external lines, will be divergent above some order of perturbation. We call such theories *non-renormalizable*.

This power-counting analysis can be repeated for any quantum field theory. As a second example, we can look at QED. We should now distinguish between photon and electron lines, which we shall denote by I_γ, $I_{\rm e}$, E_γ and $E_{\rm e}$ for internal and external lines, respectively. Taking into account the fact that the fermion propagator behaves like k^{-1} at large momenta, for the superficial degree of divergence of an 1PI diagram we obtain

$$d = 2I_\gamma + 3I_{\rm e} - 4n + 4 = 4 - E_\gamma - \frac{3}{2}E_{\rm e}\ . \tag{A8}$$

We see that d is independent of the order of perturbation theory and, therefore, the theory is renormalizable.

We leave it as an exercise to the reader to establish the renormalization properties of other field theories. In four dimensions of space–time, the result is:

1. there exists only one super-renormalizable field-theory with interaction of the form ϕ^3;
2. there exist five renormalizable ones:
   (a) ϕ^4;
   (b) Yukawa $\bar{\psi}\psi\phi$;
   (c) QED $\bar{\psi}\gamma_\mu A^\mu \psi$;
   (d) scalar electrodynamics, it contains two terms $[\phi^\dagger \partial_\mu \phi - (\partial_\mu \phi^\dagger)\phi]A^\mu$ and $A^\mu A_\mu \phi^\dagger \phi$;
   (e) Yang–Mills $\mathrm{Tr}\, G_{\mu\nu}G^{\mu\nu}$;
3. all other theories are non-renormalizable.

For ϕ^3, the energy will turn out to be unbounded from below, so this theory alone cannot be a fundamental theory for a physical system. A most remarkable fact is that, as we shall see later, nature uses *all* renormalizable theories to describe the interactions among elementary particles.

Before closing this section we want to make a remark which is based on ordinary dimensional analysis. In four dimensions, a boson field has dimensions of a mass (remember, we are using units such that the speed of light c and Planck's constant h are dimensionless) and a fermion field with a mass to the power $3/2$. Since all terms in a Lagrangian density must have dimensions equal to four, we conclude that the coupling constant of a super-renormalizable theory must have the dimensions of a mass, a renormalizable theory must be dimensionless, and a non-renormalizable theory must have the dimensions of an inverse power of mass. In fact we can rephrase the power-counting argument for the superficial degree of divergence of an 1PI diagram as an argument based on dimensional analysis. The result will be this connection between the dimensions of the coupling constant and the renormalization properties of the theory. However, there is a fine point: for this argument to work we must assume that all boson propagators behave like k^{-2} at large momenta and all fermion ones like k^{-1}. So, the argument will fail if this behaviour is not true. The most important example of such a failure is a theory containing massive vector fields whose propagator is like a constant at large k. As a result, such theories, although they may have dimensionless coupling constants, are in fact non-renormalizable.

A.2.2 Regularization

The point splitting we presented in Eq. (A4) is an example of a procedure we shall call *regularization*. It consists of introducing an extra parameter in the theory (in the case considered, it was the splitting distance a), to which we do not necessarily attach a physical meaning, with the following properties: (i) the initial theory is recovered for a particular value of the parameter, in our example $a = 0$; (ii) the theory is finite for all values of the parameter in a region which contains the 'physical' one $a = 0$; and (iii) at this value we get back the divergences of the initial theory. We shall call this parameter *a cut-off*.

If our purpose is to perform computations of Feynman diagrams, we may choose any cut-off procedure that renders these diagrams finite. There is a plethora of such methods and there is no need to give a complete list. A direct method would be to cut all integrations of loop momenta at a scale Λ. The initial theory is recovered at the limit $\Lambda \to \infty$. For practical calculations it is clear that we must choose a cut-off procedure that renders these computations as simple as possible. By trial and error, the simplest regularization scheme turned out to be a quite counter-intuitive one. We start by illustrating it in the simple example of the divergent integral of Eq. (A1). Since we are interested only in the divergent part, we can simplify the discussion by considering the value of I at $p = 0$. We thus obtain

$$I = \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{(k^2 - m^2 + \mathrm{i}\epsilon)^2} \,. \tag{A9}$$

Ignoring the divergence for the moment, we notice that the integrand depends only on k^2, so we choose spherical coordinates and write $\mathrm{d}^4k = k^3\mathrm{d}kd\Omega^{(3)}$, where $d\Omega^{(3)}$ is the surface element on the three-dimensional unit sphere. Further, we notice that I would have been convergent if we were working in a space–time of three, two or one dimensions. The crucial observation is that in all three cases we can write the result in a compact form as follows[15]:

$$I^{(d)} = \int \frac{\mathrm{d}^dk}{(2\pi)^d}\frac{1}{(k^2+m^2)^2} = \frac{1}{(4\pi)^{d/2}}\frac{\Gamma(2-d/2)}{(m^2)^{(2-d/2)}} \; ; \quad d = 1,2,3, \tag{A10}$$

where $\Gamma(z)$ is the well-known special function which generalizes the concept of the factorial for a complex z. The important values for Eq. (A10) are given by

$$\Gamma(n) = (n-1)! \; ; \quad \Gamma(n+1/2) = \frac{(\pi)^{1/2}}{2^n}(2n-1)!! \; ; \quad n = 1,2,\ldots \; . \tag{A11}$$

And now comes the big step. Nothing on the right-hand side of Eq. (A10) forces us to consider this expression only for $d = 1,2$ or 3. In fact, Γ is a meromorphic function in the entire complex plane with poles whenever its argument becomes equal to an integer $n \leq 0$. For the integral $I^{(d)}$, using the identity $n\Gamma(n) = \Gamma(n+1)$, we see that, when $d \to 4$, the Γ function behaves as $\Gamma(2-d/2) \sim 2/(4-d)$. So we can argue that, at least for this integral, we have introduced a regularization, i.e. a new parameter, namely $\epsilon = 4-d$, such that the expression is well defined for all values in a region of ϵ and diverges when $\epsilon \to 0$.

Before showing how to generalize this approach to all other integrals we may encounter in the calculation of Feynman diagrams, let us try to make the logic clear by emphasizing what this regularization does not claim to be. First, it does not claim to be the result one would have obtained by quantizing the theory in a complex number of dimensions. In fact we do not know how to consistently perform such an operation. In this sense, dimensional regularization does not offer a non-perturbative definition of the field theory. The prescription applies directly to the integrals obtained order by order in the perturbation expansion. Second, it cannot even be viewed as the analytic continuation to the complex d plane of the results we obtain in performing the integral for $d = 1,2,3$. Indeed, the knowledge of the values of a function on a finite number of points on the real axis does not allow for a unique analytic continuation. Instead, the claim is that Eq. (A10), appropriately generalized, offers an unambiguous prescription to obtain a well-defined answer for any Feynman diagram as long as ϵ stays away from zero.

The observation which allows for such a generalization is that Feynman rules always yield a special class of integrals. In purely bosonic theories, whether renormalizable or not, they are of the form

$$I(p_1,p_2,\ldots,p_n) = \int \prod_i \left(\frac{\mathrm{d}^dk_i}{(2\pi)^d}\right)\frac{N(k_1,k_2,\ldots)}{D(k_1,k_2,\ldots)}\prod_r \Big((2\pi)^d\delta^d(k,p)\Big) \; , \tag{A12}$$

where the k and the p are the momenta of the internal and external lines respectively, the product over i runs over all internal lines, that of r over all vertices, the δ functions denote the energy and momentum conservation on every vertex, and N and D are polynomials of the form

$$N(k_1,k_2,\ldots) = k_1^{\mu_1}k_1^{\mu_2}\ldots k_2^{\nu_1}k_2^{\nu_2}\ldots , \tag{A13}$$

$$D(k_1,k_2,\ldots) = \prod_i (k_i^2+m_i^2) \; . \tag{A14}$$

D is just the product of all propagators and m_i is the mass of the ith line. N appears through derivative couplings and/or the $k^\mu k^\nu$ parts of the propagators of higher-spin bosonic fields. It equals

[15] We write the result after a Wick rotation in Euclidean space

one in theories with only scalar fields and non-derivative couplings, such as ϕ^4. All scalar products are written in terms of the d-dimensional Euclidean metric $\delta_{\mu\nu}$ which satisfies

$$\delta^\mu_\mu = \mathrm{Tr}\,\mathbb{1} = d\,. \tag{A15}$$

The dimensional regularization consists of giving a precise expression for $I(p_1, p_2, \ldots, p_n)$ as a function of d which coincides with the usual value whenever the latter exists and is well defined for every value of d in the complex d plane except for those positive integer values for which the original integral is divergent.

At one loop the integral Eq. (A12) reduces to

$$I(p_1, p_2, \ldots, p_n) = \int \frac{d^d k}{(2\pi)^d} \frac{N(k)}{D(k, p_1, p_2, \ldots)}\,, \tag{A16}$$

with k being the loop momentum. The denominator D is of the form

$$D(k, p_1, p_2, \ldots) = \prod_i [(k - \Sigma_{(i)} p)^2 + m_i^2]\,, \tag{A17}$$

where $\Sigma_{(i)} p$ denotes the combination of external momenta which goes through the ith internal line. This product of propagators can be cast in a more convenient form by using a formula first introduced by Feynman:

$$\frac{1}{P_1 P_2 \ldots P_\eta} = (\eta - 1)! \int_0^1 \frac{\mathrm{d}z_1 \mathrm{d}z_2 \ldots \mathrm{d}z_\eta \delta(1 - \Sigma_i z_i)}{[z_1 P_1 + z_2 P_2 + \cdots + z_\eta P_\eta]^\eta}\,. \tag{A18}$$

With the help of Eq. (A18) and an appropriate change of variables, all one-loop integrals become of the general form

$$\hat{I}(p_1, p_2, \ldots, p_n) = \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{k_{\mu_1} k_{\mu_2} \ldots k_{\mu_l}}{[k^2 + F^2(p, m, z)]^\eta}\,, \tag{A19}$$

with F some scalar function of the external momenta, the masses and the Feynman parameters. F has the dimensions of a mass. $I(p_1, p_2, \ldots, p_n)$ is obtained from $\hat{I}(p_1, p_2, \ldots, p_n)$ after integration with respect to the Feynman parameters z_i of Eq. (A18). For odd values of l, $\hat{I}$ vanishes by symmetric integration. For l even it can be easily computed using spherical coordinates. Some simple cases are as follows:

$$\int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{1}{[k^2 + F^2(p, m, z)]^\eta} = \frac{1}{(4\pi)^{d/2}} \frac{\Gamma(\eta - d/2)}{\Gamma(\eta)} [F^2]^{(d/2 - \eta)}\,. \tag{A20}$$

$$\int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{k_\mu k_\nu}{[k^2 + F^2(p, m, z)]^\eta} = \frac{1}{(4\pi)^{d/2}} \frac{\delta_{\mu\nu}}{2} \frac{\Gamma(\eta - 1 - d/2)}{\Gamma(\eta)} [F^2]^{(d/2 + 1 - \eta)}\,. \tag{A21}$$

At the end, we are interested in the limit $d \to 4$. The first integral Eq. (A20) diverges for $\eta \leq 2$ and the second Eq. (A21) for $\eta \leq 3$. For $\eta = 2$ and $d = 4$, Eq. (A20) is logarithmically divergent and our regularized expression is regular for $\mathrm{Re}\, d < 4$ and presents a simple pole $\sim 1/(d - 4)$. For $\eta = 1$, it is quadratically divergent but our expression still has a simple pole at $d = 4$. The difference is that now the first pole from the left is at $d = 2$. We arrive at the same conclusions looking at the integral of Eq. (A21): by dimensionally regularizing a one-loop integral corresponding to a Feynman diagram which, by power counting, diverges as Λ^{2n}, we obtain a meromorphic function of d with simple poles starting at $d = 4 - 2n$. By convention, $n = 0$ denotes a logarithmic divergence.

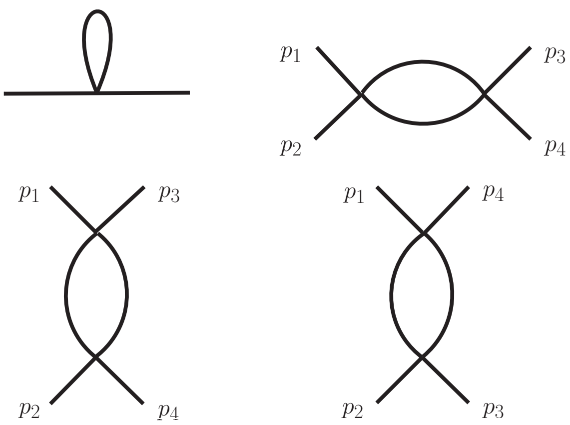

Fig. A.3: The one loop primitively divergent diagrams of the ϕ^4 theory

A.2.3 Renormalization

In this section, we want to address the physical question of under which circumstances can a meaningful four-dimensional theory be recovered from the regularized ϵ-dependent expressions. As one could have anticipated, the answer will turn out to be that this is only possible for the renormalizable (and super-renormalizable) theories we introduced before. The procedure to do so is called *renormalization*. In this section, we shall present some simple examples.

Let us start with the simplest four-dimensional renormalizable theory given by our already familiar Lagrangian density from Eq. (A2). In $d = 4$, the field ϕ has the dimensions of a mass and the coupling constant λ is dimensionless. Since we intend to use dimensional regularization, we introduce a mass parameter μ and write the coefficient of the interaction term $\lambda \to \mu^\epsilon \lambda$, so that the coupling constant λ remains dimensionless at all values of ϵ. We shall present the renormalization programme for this theory at the lowest non-trivial order, that which includes all diagrams up to and including those with one closed loop.

The power-counting argument presented previously shows that, at one loop, the only divergent 1PI diagrams are the ones of Fig. A.3.

The two-point diagram is quadratically divergent and the four-point diagram is logarithmically divergent[16]. We choose to work entirely with dimensional regularization and for these diagrams in Minkowski space–time, using (A20) at the limit $d \to 4$, we obtain

$$\Gamma_1^{(2)} = \frac{\lambda\mu^\epsilon}{2}\int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{k^2 - m^2} = \frac{\mathrm{i}\lambda m^2}{16\pi^2}\,\frac{1}{\epsilon} \tag{A22}$$

$$\begin{aligned}\Gamma_1^{(4)}(p_1,\ldots,p_4) &= \frac{1}{2}\lambda^2\mu^{2\epsilon}\int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{(k^2-m^2)[(k-P)^2-m^2]} + \text{crossed}\\ &= \frac{1}{2}\lambda^2\mu^{2\epsilon}\int_0^1 dz\int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{[k^2-m^2+P^2 z(1-z)]^2} + \text{crossed}\\ &= \frac{3\mathrm{i}\lambda^2}{16\pi^2}\,\frac{1}{\epsilon} + \text{finite terms}\,,\end{aligned} \tag{A23}$$

[16] We could prevent the appearance of the first diagram by 'normal ordering' the ϕ^4 term in the interaction Lagrangian, but, for pedagogical purposes, we prefer not to do so. Normal ordering is just a particular prescription to avoid certain divergences, but it is not always the most convenient one. First, it is not general. For example, it will not prevent the appearance of divergence in the two-point function at higher orders and second, its use may complicate the discussion of possible gauge symmetries of $\mathcal{L}$.

where $P = p_1 + p_2$, 'crossed' stands for the contribution of the two crossed diagrams in Fig. A.3 and 'finite terms' represent the contributions which are regular when $d = 4$. We can make the following remarks.

1. The divergent contributions are constants, independent of the external momenta. We shall see shortly, in the example of QED, that this is a particular feature of the ϕ^4 theory. In fact, even for ϕ^4, it is no longer true when higher loops are considered. However, we can prove the following general property: all divergent terms are proportional to monomials in the external momenta. We have already introduced this result. For one-loop diagrams the proof is straightforward. We start from the general expression of Eq. (A19) and notice that we can expand the integrand in powers of the external momenta p taken around some fixed point. Every term in this expansion increases the value of η, so after a finite number of terms, the integral becomes convergent. It takes some more work to generalize the proof to multi-loop diagrams, but it can be done,
2. The dependence of the divergent terms on m^2 could be guessed from dimensional analysis. This is one of the attractive features of dimensional regularization,
3. The finite terms in Eq. (A23) depend on the parameter μ. The Laurent expansion in ϵ brings terms of the form ln$\{[m^2 - P^2 z(1-z)]/\mu^2\}$.

The particular form of the divergent terms suggests the prescription to remove them. Let us start with the two-point function. In the loop expansion we write

$$\Gamma^{(2)}(p^2) = \sum_{l=0}^{\infty} \Gamma_l^{(2)}(p^2) = \Gamma_0^{(2)}(p^2) + \Gamma_1^{(2)}(p^2) + \cdots \ , \tag{A24}$$

where the index l denotes the contribution of the diagrams with l loops. In the tree approximation we have

$$\Gamma_0^{(2)}(p^2) = -\mathrm{i}(p^2 - m^2) \ . \tag{A25}$$

The one-loop diagram adds the term given by Eq. (A22). Since it is a constant, it can be interpreted as a correction to the value of the mass in Eq. (A25). Therefore, we can introduce a *renormalized* mass m_{R}^2, which is a function of m, λ and ϵ. Of course, this function can only be computed as a formal power series in λ. Up to and including one-loop diagrams we write

$$m_{\mathrm{R}}^2(m, \lambda, \epsilon) = m^2 \left(1 + \frac{\lambda}{16\pi^2} \frac{1}{\epsilon}\right) + \mathrm{O}(\lambda^2) \ . \tag{A26}$$

A formal power series whose zero-order term is non-vanishing is invertible in terms of another formal power series. So, we can write m as a function of m_{R}, λ and ϵ:

$$m^2(m_{\mathrm{R}}, \lambda, \epsilon) = m_{\mathrm{R}}^2 \left(1 - \frac{\lambda}{16\pi^2} \frac{1}{\epsilon}\right) + \mathrm{O}(\lambda^2) \equiv m_{\mathrm{R}}^2 Z_m + \mathrm{O}(\lambda^2) \ , \tag{A27}$$

where we have defined the function $Z_m(\lambda, \epsilon)$ as a formal power series in λ with ϵ-dependent coefficients.

The parameter m is often called the *bare* mass. In the Lagrangian Eq. (A2), replacing the bare mass m with the help of Eq. (A27) results in: (i) changing the Feynman rules m by m_{R} and (ii) introducing a new term in $\mathcal{L}$ of the form

$$\delta\mathcal{L}_m = m_{\mathrm{R}}^2 \frac{\lambda}{32\pi^2} \frac{1}{\epsilon} \phi^2(x) \ . \tag{A28}$$

Fig. A.4: The new diagram resulting from $\delta\mathcal{L}_m$ of Eq. (A28)

Since $\delta\mathcal{L}_m$ is proportional to the coupling constant λ, we can view it as a new vertex in the perturbation expansion which, to first order, gives the diagram of Fig. A.4. In this case the complete two-point function to first order in λ is given by

$$\begin{aligned}\Gamma^{(2)}(p^2) &= -\mathrm{i}(p^2 - m_\mathrm{R}^2) + \frac{\mathrm{i}\lambda m_\mathrm{R}^2}{16\pi^2}\,\frac{1}{\epsilon} - \frac{\mathrm{i}\lambda m_\mathrm{R}^2}{16\pi^2}\,\frac{1}{\epsilon} + \mathrm{O}(\lambda^2) \\ &= -\mathrm{i}(p^2 - m_\mathrm{R}^2) + \mathrm{O}(\lambda^2)\ ,\end{aligned} \tag{A29}$$

which means that, if we keep fixed m_R and λ instead of m and λ, we can take the limit $d \to 4$ and find no divergences up to and including one-loop diagrams for the two-point function.

Now that we have understood the principle, it is straightforward to apply it to the four-point function. In the same spirit we write

$$\Gamma^{(4)}(p_1, \ldots, p_4) = \sum_{l=0}^{\infty} \Gamma_l^{(4)}(p_1, \ldots, p_4) = \Gamma_0^{(4)}(p_1, \ldots, p_4) + \Gamma_1^{(4)}(p_1, \ldots, p_4) + \cdots\ . \tag{A30}$$

In the tree approximation, $\Gamma_0^{(4)}(p_1, \ldots, p_4) = -\mathrm{i}\lambda$. Including the one-loop diagrams we obtain

$$\Gamma^{(4)}(p_1, \ldots, p_4) = -\mathrm{i}\lambda\left(1 - \frac{3\lambda}{16\pi^2}\,\frac{1}{\epsilon} + \text{finite terms}\right) + \mathrm{O}(\lambda^3)\ . \tag{A31}$$

We change from the *bare* coupling constant λ to the *renormalized* one λ_R by writing

$$\lambda_\mathrm{R}(\lambda, \epsilon) = \lambda\left(1 - \frac{3\lambda}{16\pi^2}\,\frac{1}{\epsilon} + \mathrm{O}(\lambda^2)\right)\ , \tag{A32}$$

or, equivalently,

$$\lambda(\lambda_\mathrm{R}, \epsilon) = \lambda_\mathrm{R}\left(1 + \frac{3\lambda_\mathrm{R}}{16\pi^2}\,\frac{1}{\epsilon} + \mathrm{O}(\lambda_\mathrm{R}^2)\right) \equiv \lambda_\mathrm{R} Z_\lambda\ . \tag{A33}$$

Again, replacing λ with λ_R in $\mathcal{L}$ produces a new four-point vertex which cancels the divergent part of the one-loop diagrams of Fig. A.3. Let us also notice that we can replace λ with λ_R in Eq. (A27) since the difference will appear only at the higher order.

Until now we have succeeded in building a new, *renormalized* Lagrangian, and the resulting theory is free from divergences up to and including one-loop diagrams. It involves two new terms which change the coefficients of the ϕ^2 and ϕ^4 terms of the original Lagrangian. These terms are usually called *counter-terms*. They are the expression, in terms of the dimensional regularization cut-off parameter ϵ, of the process we outlined in Eqs. (A4) and (A5). They provide the correct definition, up to this order of perturbation, of the Lagrangian density, by removing the short-distance ambiguities inherent in the local expressions ϕ^2 and ϕ^4.

Before looking at higher orders, let us see the price we had to pay for this achievement. It can be better seen at the four-point function. Looking back at the Eq. (A23), we make the following two observations. First, as we noticed already, the finite part seems to depend on a new arbitrary parameter with the dimensions of a mass μ. Second, the definition of Z_λ in Eq. (A33) also seems arbitrary. We

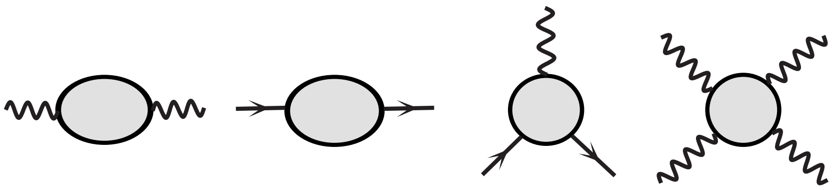

Fig. A.5: The primitively divergent 1PI Green's functions of QED. The last one, the light-by-light scattering, is convergent as a consequence of gauge invariance.

could add to it any term of the form $C\lambda_{\mathrm{R}}$ with C any arbitrary constant independent of ϵ. Such an addition would change the value of the coupling constant at the one-loop order. The two observations are not unrelated. Indeed, changing the parameter μ from μ_1 to μ_2 in Eq. (A23) adds a constant term proportional to $\lambda\ \ln(\mu_1/\mu_2)$ which, as we just saw, can be absorbed in a redefinition of Z_λ and thus of the value of the coupling constant. This μ dependance can be studied systematically and gives rise to the renormalization group equation which I will not present here. We conclude that, at the one-loop level, all arbitrariness of the renormalization programme consists of assigning prescribed values to two parameters of the theory, which can be chosen to be the mass and the coupling constant. A convenient choice is given by two conditions of the form

$$\Gamma^{(2)}(p^2 = m_{\mathrm{R}}^2) = 0 \tag{A34}$$

and

$$\Gamma^{(4)}(p_1, \ldots, p_4)|_{\text{point } M} = \mathrm{i}\lambda_{\mathrm{R}}^{(M)}\ . \tag{A35}$$

The first one, Eq. (A34), defines the physical mass as the pole of the complete propagator. Although this choice is the most natural for physics, from a purely technical point of view, we could use any condition assigning a prescribed value to $\Gamma^{(2)}(p^2)$ at a fixed point $p^2 = M^2$, provided it is a point in which $\Gamma^{(2)}(p^2)$ is regular. Similarly, in the second condition Eq. (A35), by 'point M' we mean some point in the space of the four momenta p_i, $i = 1, \ldots, 4$, provided it is a point in which $\Gamma^{(4)}$ is regular. For a massive theory the point $p_i = 0$ is an example. Once these conditions are imposed, all Green functions at one loop are well defined and calculable. A final remark: at one loop no counter-term corresponding to the kinetic energy term $(\partial_\mu\phi)^2$ is needed. This is an accident of the one-loop for the ϕ^4 theory. It appears only at higher orders.

This process of removing the ambiguities by introducing counter-terms in the original Lagrangian can be extended to all orders of perturbation. The proof is rather complicated but essentially elementary. No new ideas are necessary. We must prove that, at any order, the terms appear with the correct combinatoric factor, even in the cases in which sub-diagrams are divergent to which counter-terms have already been assigned. At the end, all Green functions of a renormalizable theory, or any combination of renormalizable theories, are well defined and calculable.

As a second example, we shall present the renormalization for the one-loop diagrams of QED. The method is exactly the same and yields 'renormalized' values of the various terms which appear in the QED Lagrangian. Looking at the power-counting Eq. (A8), we see that the only possibly divergent 1PI diagrams with one loop are those of Fig. A.5. A simple calculation gives:

– the photon self-energy

$$\Gamma_{\mu\nu}^{(2,0)}(q) = \frac{2\mathrm{i}\alpha}{3\pi}\frac{1}{\epsilon}(q_\mu q_\nu - q^2 g_{\mu\nu}) + \cdots\ , \tag{A36}$$

where $\alpha = e^2/4\pi$ is the fine-structure constant and the dots stand for finite terms;

– the electron self-energy

$$\Gamma^{(0,2)}(p) = \frac{\mathrm{i}\alpha}{2\pi}\frac{1}{\epsilon}\not{p} - \frac{2\mathrm{i}\alpha}{\pi}\frac{1}{\epsilon}m + \cdots \ , \tag{A37}$$

where we have suppressed spinor indices and, again, the dots stand for finite terms—we can see that in Eqs. (A36) and (A37) the divergent terms are monomials in the external momenta;

– the vertex function

$$\Gamma^{(1,2)}_{\mu}(p,p') = \frac{\mathrm{i}\alpha}{2\pi}\frac{1}{\epsilon}e\gamma_{\mu} + \cdots \ . \tag{A38}$$

As before, all these divergences can be absorbed in the definition of renormalized quantities as

$$A^{\mu}(x) = Z_3^{1/2}A^{\mu}_{\mathrm{R}}(x) = \left(1 - \frac{\alpha}{3\pi}\frac{1}{\epsilon} + \mathrm{O}(\alpha^2)\right)A^{\mu}_{\mathrm{R}}(x), \tag{A39}$$

$$\psi(x) = Z_2^{1/2}\psi_{\mathrm{R}}(x) = \left(1 - \frac{\alpha}{4\pi}\frac{1}{\epsilon} + \mathrm{O}(\alpha^2)\right)\psi_{\mathrm{R}}(x), \tag{A40}$$

$$m = Z_m m_{\mathrm{R}} = \left(1 - \frac{2\alpha}{\pi}\frac{1}{\epsilon} + \mathrm{O}(\alpha^2)\right)m_{\mathrm{R}}, \tag{A41}$$

$$\Gamma^{(1,2)}_{\mu}(p,p') = -\mathrm{i}eZ_1\gamma_{\mu} + \cdots = -\mathrm{i}e\gamma_{\mu}\left(1 - \frac{\alpha}{2\pi}\frac{1}{\epsilon} + \mathrm{O}(\alpha^2)\right) + \cdots \ . \tag{A42}$$

As we noticed already, in QED the counter-terms corresponding to the kinetic energies of the electron and the photon appear already at the one-loop order. Putting all counter-terms together, the interaction Lagrangian becomes:

$$-e\bar{\psi}\gamma_{\mu}\psi A^{\mu} = -Z_e Z_2 Z_3^{1/2} e_{\mathrm{R}}\bar{\psi}_{\mathrm{R}}\gamma_{\mu}\psi_{\mathrm{R}}A^{\mu}_{\mathrm{R}} \ . \tag{A43}$$

It follows that the condition which determines the charge renormalization constant Z_e is

$$Z_e Z_2 Z_3 = Z_1 \ . \tag{A44}$$

By comparing Eqs. (A42) and (A40), we see that, at least at this order, $Z_1 = Z_2$. Therefore, the entire charge renormalization is determined by the photon self-energy diagram. We can show that this property is valid to all orders of perturbation theory and is a consequence of gauge invariance. It is the same property of gauge invariance which guarantees that the last diagram of Fig. A.5, when computed using dimensional regularization which respects gauge invariance, is in fact finite.

This completes a very sketchy discussion of renormalization theory. Only straightforward calculations are needed to adapt it to any renormalizable theory and to any order in the perturbation expansion.