Published by CERN in the Proceedings of the 2013 CERN–Latin-American School of High-Energy Physics, Arequipa, Peru, 6 – 19 March 2013, edited by M. Mulders and G. Perez, CERN-2015-001 (CERN, Geneva, 2015)

Introductory Lectures on Quantum Field Theory

L. Álvarez-Gaumé [a] *and M. A. Vázquez-Mozo* [b]
[a] CERN, Geneva, Switzerland
[b] Universidad de Salamanca, Salamanca, Spain

Abstract
In these lectures we present a few topics in quantum field theory in detail. Some of them are conceptual and some more practical. They have been selected because they appear frequently in current applications to particle physics and string theory.

1 Introduction

These notes summarize lectures presented at the 2005 CERN-CLAF School in Malargüe (Argentina), the 2009 CERN-CLAF School in Medellín (Colombia), the 2011 CERN-CLAF School in Natal (Brazil), the 2012 Asia-Europe-Pacific School of High Energy Physics in Fukuoka (Japan), and the 2013 CERN–Latin-American School of High-Energy Physics in Arequipa (Peru). The audience in all occasions was composed to a large extent by students in experimental High Energy Physics with an important minority of theorists. In nearly ten hours it is quite difficult to give a reasonable introduction to a subject as vast as quantum field theory. For this reason the lectures were intended to provide a review of those parts of the subject to be used later by other lecturers. Although a cursory acquaintance with th subject of quantum field theory is helpful, the only requirement to follow the lectures it is a working knowledge of Quantum Mechanics and Special Relativity.

The guiding principle in choosing the topics presented (apart to serve as introductions to later courses) was to present some basic aspects of the theory that present conceptual subtleties. Those topics one often is uncomfortable with after a first introduction to the subject. Among them we have selected:

- The need to introduce quantum fields, with the great complexity this implies.
- Quantization of gauge theories and the rôle of topology in quantum phenomena. We have included a brief study of the Aharonov-Bohm effect and Dirac's explanation of the quantization of the electric charge in terms of magnetic monopoles.
- Quantum aspects of global and gauge symmetries and their breaking.
- Anomalies.
- The physical idea behind the process of renormalization of quantum field theories.
- Some more specialized topics, like the creation of particle by classical fields and the very basics of supersymmetry.

These notes have been written following closely the original presentation, with numerous clarifications. Sometimes the treatment given to some subjects has been extended, in particular the discussion of the Casimir effect and particle creation by classical backgrounds. Since no group theory was assumed, we have included an Appendix with a review of the basics concepts.

By lack of space and purpose, few proofs have been included. Instead, very often we illustrate a concept or property by describing a physical situation where it arises. A very much expanded version of these lectures, following the same philosophy but including many other topics, has appeared in book form in [1]. For full details and proofs we refer the reader to the many textbooks in the subject, and in particular in the ones provided in the bibliography [2–11]. Specially modern presentations, very much in the spirit of these lectures, can be found in references [5, 6, 10, 11]. We should nevertheless warn the reader that we have been a bit cavalier about references. Our aim has been to provide mostly a (not exhaustive) list of reference for further reading. We apologize to those authors who feel misrepresented.

978–92–9083–412-0; 0531-4283 –
http://dx.doi.org/10.5170/CERN-2015-001.1

A note about notation

Before starting it is convenient to review the notation used. Through these notes we will be using the metric $\eta_{\mu\nu} = \mathrm{diag}\,(1,-1,-1,-1)$. Derivatives with respect to the four-vector $x^\mu = (ct, \vec{x})$ will be denoted by the shorthand

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\frac{1}{c}\frac{\partial}{\partial t}, \vec{\nabla}\right). \tag{1}$$

As usual space-time indices will be labelled by Greek letters ($\mu, \nu, \ldots = 0, 1, 2, 3$) while Latin indices will be used for spatial directions ($i, j, \ldots = 1, 2, 3$). In many expressions we will use the notation $\sigma^\mu = (\mathbf{1}, \sigma^i)$ where σ^i are the Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

Sometimes we use of the Feynman's slash notation $\not{a} = \gamma^\mu a_\mu$. Finally, unless stated otherwise, we work in natural units $\hbar = c = 1$.

2 Why do we need quantum field theory after all?

In spite of the impressive success of Quantum Mechanics in describing atomic physics, it was immediately clear after its formulation that its relativistic extension was not free of difficulties. These problems were clear already to Schrödinger, whose first guess for a wave equation of a free relativistic particle was the Klein-Gordon equation

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 + m^2\right)\psi(t, \vec{x}) = 0. \tag{3}$$

This equation follows directly from the relativistic "mass-shell" identity $E^2 = \vec{p}^2 + m^2$ using the correspondence principle

$$\begin{aligned} E &\rightarrow i\frac{\partial}{\partial t}, \\ \vec{p} &\rightarrow -i\vec{\nabla}. \end{aligned} \tag{4}$$

Plane wave solutions to the wave equation (3) are readily obtained

$$\psi(t, \vec{x}) = e^{-ip_\mu x^\mu} = e^{-iEt + i\vec{p}\cdot\vec{x}} \qquad \text{with} \qquad E = \pm\omega_p \equiv \pm\sqrt{\vec{p}^2 + m^2}. \tag{5}$$

In order to have a complete basis of functions, one must include plane wave with both $E > 0$ and $E < 0$. This implies that given the conserved current

$$j_\mu = \frac{i}{2}\Big(\psi^*\partial_\mu\psi - \partial_\mu\psi^*\,\psi\Big), \tag{6}$$

its time-component is $j^0 = E$ and therefore does not define a positive-definite probability density.

A complete, properly normalized, continuous basis of solutions of the Klein-Gordon equation (3) labelled by the momentum $\vec{p}$ can be defined as

$$\begin{aligned} f_p(t, \vec{x}) &= \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{2\omega_p}}\, e^{-i\omega_p t + i\vec{p}\cdot\vec{x}}, \\ f_{-p}(t, \vec{x}) &= \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{2\omega_p}}\, e^{i\omega_p t - i\vec{p}\cdot\vec{x}}. \end{aligned} \tag{7}$$

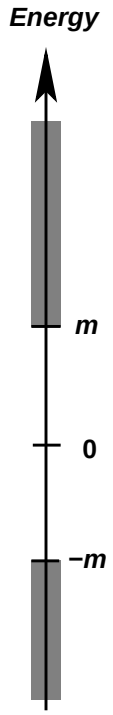


Fig. 1: Spectrum of the Klein-Gordon wave equation

Given the inner product

$$\langle\psi_1|\psi_2\rangle = i\int d^3x\Big(\psi_1^*\partial_0\psi_2 - \partial_0\psi_1^*\,\psi_2\Big)$$

the states (7) form an orthonormal basis

$$\begin{aligned}\langle f_p|f_{p'}\rangle &= \delta(\vec{p}-\vec{p}\,'),\\ \langle f_{-p}|f_{-p'}\rangle &= -\delta(\vec{p}-\vec{p}\,'), \qquad (8)\\ \langle f_p|f_{-p'}\rangle &= 0. \qquad (9)\end{aligned}$$

The wave functions $f_p(t,x)$ describes states with momentum $\vec{p}$ and energy given by $\omega_p = \sqrt{\vec{p}^{\,2}+m^2}$. On the other hand, the states $|f_{-p}\rangle$ not only have a negative scalar product but they actually correspond to negative energy states

$$i\partial_0 f_{-p}(t,\vec{x}) = -\sqrt{\vec{p}^{\,2}+m^2}\, f_{-p}(t,\vec{x}). \qquad (10)$$

Therefore the energy spectrum of the theory satisfies $|E| > m$ and is unbounded from below (see Fig. 1). Although in a case of a free theory the absence of a ground state is not necessarily a fatal problem, once the theory is coupled to the electromagnetic field this is the source of all kinds of disasters, since nothing can prevent the decay of any state by emission of electromagnetic radiation.

The problem of the instability of the "first-quantized" relativistic wave equation can be heuristically tackled in the case of spin-$\frac{1}{2}$ particles, described by the Dirac equation

$$\left(-i\beta\frac{\partial}{\partial t} + \vec{\alpha}\cdot\vec{\nabla} - m\right)\psi(t,\vec{x}) = 0, \qquad (11)$$

where $\vec{\alpha}$ and β are 4×4 matrices

$$\alpha^i = \begin{pmatrix} 0 & i\sigma^i \\ -i\sigma^i & 0 \end{pmatrix}, \qquad \beta = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \qquad (12)$$

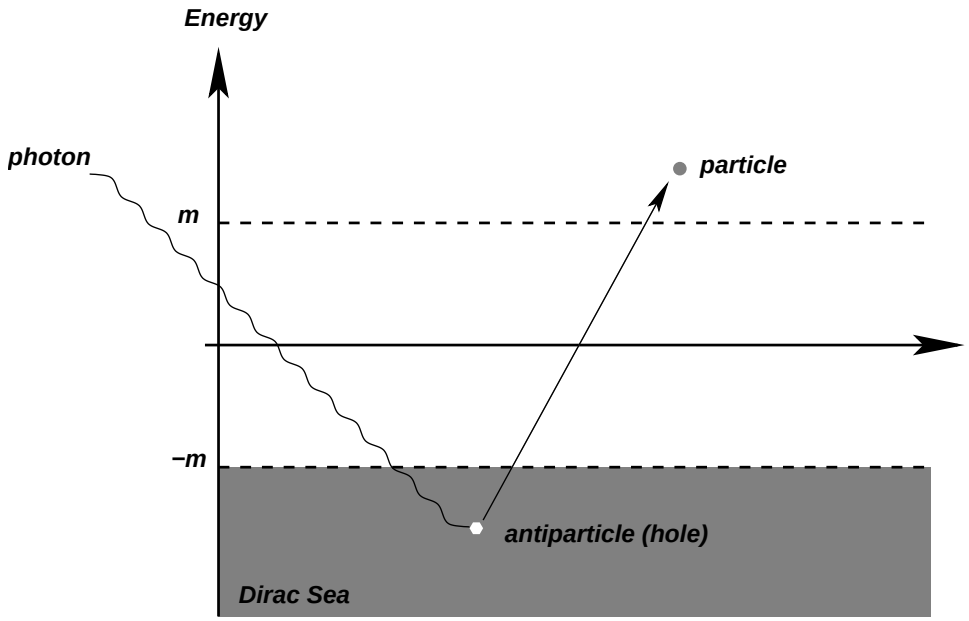


Fig. 2: Creation of a particle-antiparticle pair in the Dirac see picture

with σ^i the Pauli matrices, and the wave function $\psi(t,\vec{x})$ has four components. The wave equation (11) can be thought of as a kind of "square root" of the Klein-Gordon equation (3), since the latter can be obtained as

$$\left(-i\beta\frac{\partial}{\partial t}+\vec{\alpha}\cdot\vec{\nabla}-m\right)^{\dagger}\left(-i\beta\frac{\partial}{\partial t}+\vec{\alpha}\cdot\vec{\nabla}-m\right)\psi(t,\vec{x})=\left(\frac{\partial^2}{\partial t^2}-\nabla^2+m^2\right)\psi(t,\vec{x}). \qquad (13)$$

An analysis of Eq. (11) along the lines of the one presented above for the Klein-Gordon equation leads again to the existence of negative energy states and a spectrum unbounded from below as in Fig. 1. Dirac, however, solved the instability problem by pointing out that now the particles are fermions and therefore they are subject to Pauli's exclusion principle. Hence, each state in the spectrum can be occupied by at most one particle, so the states with $E = m$ can be made stable if we assume that *all* the negative energy states are filled.

If Dirac's idea restores the stability of the spectrum by introducing a stable vacuum where all negative energy states are occupied, the so-called Dirac sea, it also leads directly to the conclusion that a single-particle interpretation of the Dirac equation is not possible. Indeed, a photon with enough energy ($E > 2m$) can excite one of the electrons filling the negative energy states, leaving behind a "hole" in the Dirac see (see Fig. 2). This hole behaves as a particle with equal mass and opposite charge that is interpreted as a positron, so there is no escape to the conclusion that interactions will produce pairs particle-antiparticle out of the vacuum.

In spite of the success of the heuristic interpretation of negative energy states in the Dirac equation this is not the end of the story. In 1929 Oskar Klein stumbled into an apparent paradox when trying to describe the scattering of a relativistic electron by a square potential using Dirac's wave equation [12] (for pedagogical reviews see [13, 14]). In order to capture the essence of the problem without entering into unnecessary complication we will study Klein's paradox in the context of the Klein-Gordon equation.

Let us consider a square potential with height $V_0 > 0$ of the type showed in Fig. 3. A solution to the wave equation in regions I and II is given by

$$\begin{aligned}\psi_I(t,x) &= e^{-iEt+ip_1x}+Re^{-iEt-ip_1x},\\ \psi_{II}(t,x) &= Te^{-iEt+p_2x}, \end{aligned} \qquad (14)$$

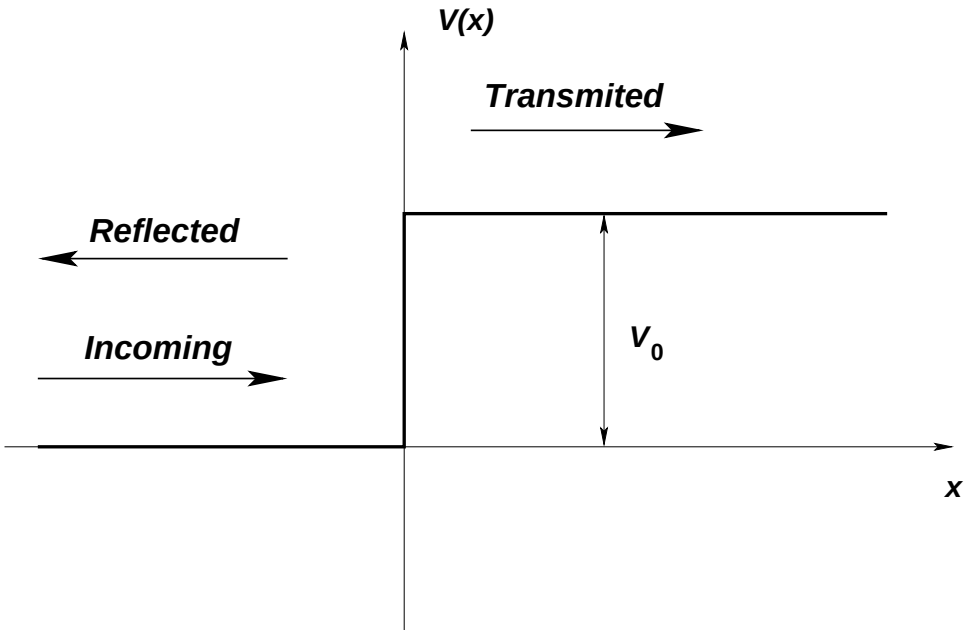


Fig. 3: Illustration of the Klein paradox.

where the mass-shell condition implies that

$$p_1 = \sqrt{E^2 - m^2}, \qquad p_2 = \sqrt{(E - V_0)^2 - m^2}. \tag{15}$$

The constants R and T are computed by matching the two solutions across the boundary $x = 0$. The conditions $\psi_I(t, 0) = \psi_{II}(t, 0)$ and $\partial_x \psi_I(t, 0) = \partial_x \psi_{II}(t, 0)$ imply that

$$T = \frac{2p_1}{p_1 + p_2}, \qquad R = \frac{p_1 - p_2}{p_1 + p_2}. \tag{16}$$

At first sight one would expect a behavior similar to the one encountered in the nonrelativistic case. If the kinetic energy is bigger than V_0 both a transmitted and reflected wave are expected, whereas when the kinetic energy is smaller than V_0 one only expect to find a reflected wave, the transmitted wave being exponentially damped within a distance of a Compton wavelength inside the barrier.

Indeed this is what happens if $E - m > V_0$. In this case both p_1 and p_2 are real and we have a partly reflected, and a partly transmitted wave. In the same way, if $V_0 - 2m < E - m < V_0$ then p_2 is imaginary and there is total reflection.

However, in the case when $V_0 > 2m$ and the energy is in the range $0 < E - m < V_0 - 2m$ a completely different situation arises. In this case one finds that both p_1 and p_2 are real and therefore the incoming wave function is partially reflected and partially transmitted across the barrier. This is a shocking result, since it implies that there is a nonvanishing probability of finding the particle at any point across the barrier with negative kinetic energy ($E - m - V_0 < 0$)! This weird result is known as Klein's paradox.

As with the negative energy states, the Klein paradox results from our insistence in giving a single-particle interpretation to the relativistic wave function. Actually, a multiparticle analysis of the paradox [13] shows that what happens when $0 < E - m < V_0 - 2m$ is that the reflection of the incoming particle by the barrier is accompanied by the creation of pairs particle-antiparticle out of the energy of the barrier (notice that for this to happen it is required that $V_0 > 2m$, the threshold for the creation of a particle-antiparticle pair).

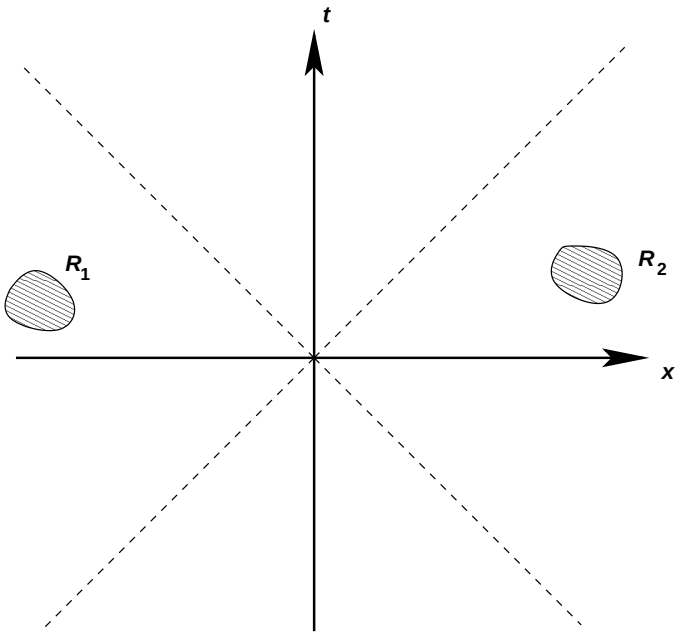


Fig. 4: Two regions R_1, R_2 that are causally disconnected.

Actually, this particle creation can be understood by noticing that the sudden potential step in Fig. 3 localizes the incoming particle with mass m in distances smaller than its Compton wavelength $\lambda = \frac{1}{m}$. This can be seen by replacing the square potential by another one where the potential varies smoothly from 0 to $V_0 > 2m$ in distances scales larger than $1/m$. This case was worked out by Sauter shortly after Klein pointed out the paradox [15]. He considered a situation where the regions with $V = 0$ and $V = V_0$ are connected by a region of length d with a linear potential $V(x) = \frac{V_0 x}{d}$. When $d > \frac{1}{m}$ he found that the transmission coefficient is exponentially small[1].

The creation of particles is impossible to avoid whenever one tries to locate a particle of mass m within its Compton wavelength. Indeed, from Heisenberg uncertainty relation we find that if $\Delta x \sim \frac{1}{m}$, the fluctuations in the momentum will be of order $\Delta p \sim m$ and fluctuations in the energy of order

$$\Delta E \sim m \tag{17}$$

can be expected. Therefore, in a relativistic theory, the fluctuations of the energy are enough to allow the creation of particles out of the vacuum. In the case of a spin-$\frac{1}{2}$ particle, the Dirac sea picture shows clearly how, when the energy fluctuations are of order m, electrons from the Dirac sea can be excited to positive energy states, thus creating electron-positron pairs.

It is possible to see how the multiparticle interpretation is forced upon us by relativistic invariance. In non-relativistic Quantum Mechanics observables are represented by self-adjoint operator that in the Heisenberg picture depend on time. Therefore measurements are localized in time but are global in space. The situation is radically different in the relativistic case. Because no signal can propagate faster than the speed of light, measurements have to be localized both in time and space. Causality demands then that two measurements carried out in causally-disconnected regions of space-time cannot interfere with each other. In mathematical terms this means that if $\mathcal{O}_{R_1}$ and $\mathcal{O}_{R_2}$ are the observables associated with two measurements localized in two causally-disconnected regions R_1, R_2 (see Fig. 4), they satisfy

$$[\mathcal{O}_{R_1}, \mathcal{O}_{R_2}] = 0, \qquad \text{if } (x_1 - x_2)^2 < 0, \text{ for all } x_1 \in R_1, x_2 \in R_2. \tag{18}$$

[1] In section (9.1) we will see how, in the case of the Dirac field, this exponential behavior can be associated with the creation of electron-positron pairs due to a constant electric field (Schwinger effect).

Hence, in a relativistic theory, the basic operators in the Heisenberg picture must depend on the space-time position x^μ. Unlike the case in non-relativistic quantum mechanics, here the position $\vec{x}$ *is not* an observable, but just a label, similarly to the case of time in ordinary quantum mechanics. Causality is then imposed microscopically by requiring

$$[\mathcal{O}(x), \mathcal{O}(y)] = 0, \qquad \text{if } (x-y)^2 < 0. \tag{19}$$

A smeared operator $\mathcal{O}_R$ over a space-time region R can then be defined as

$$\mathcal{O}_R = \int d^4x \, \mathcal{O}(x) \, f_R(x) \tag{20}$$

where $f_R(x)$ is the characteristic function associated with R,

$$f_R(x) = \begin{cases} 1 & x \in R \\ 0 & x \notin R \end{cases} . \tag{21}$$

Eq. (18) follows now from the microcausality condition (19).

Therefore, relativistic invariance forces the introduction of quantum fields. It is only when we insist in keeping a single-particle interpretation that we crash against causality violations. To illustrate the point, let us consider a single particle wave function $\psi(t, \vec{x})$ that initially is localized in the position $\vec{x} = 0$

$$\psi(0, \vec{x}) = \delta(\vec{x}). \tag{22}$$

Evolving this wave function using the Hamiltonian $H = \sqrt{-\nabla^2 + m^2}$ we find that the wave function can be written as

$$\psi(t, \vec{x}) = e^{-it\sqrt{-\nabla^2+m^2}} \delta(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x} - it\sqrt{k^2+m^2}}. \tag{23}$$

Integrating over the angular variables, the wave function can be recast in the form

$$\psi(t, \vec{x}) = \frac{1}{2\pi^2 |\vec{x}|} \int_{-\infty}^{\infty} k \, dk \, e^{ik|\vec{x}|} \, e^{-it\sqrt{k^2+m^2}}. \tag{24}$$

The resulting integral can be evaluated using the complex integration contour C shown in Fig. 5. The result is that, for any $t > 0$, one finds that $\psi(t, \vec{x}) \neq 0$ for any $\vec{x}$. If we insist in interpreting the wave function $\psi(t, \vec{x})$ as the probability density of finding the particle at the location $\vec{x}$ in the time t we find that the probability leaks out of the light cone, thus violating causality.

3 From classical to quantum fields

We have learned how the consistency of quantum mechanics with special relativity forces us to abandon the single-particle interpretation of the wave function. Instead we have to consider quantum fields whose elementary excitations are associated with particle states, as we will see below.

In any scattering experiment, the only information available to us is the set of quantum number associated with the set of free particles in the initial and final states. Ignoring for the moment other quantum numbers like spin and flavor, one-particle states are labelled by the three-momentum $\vec{p}$ and span the single-particle Hilbert space $\mathcal{H}_1$

$$|\vec{p}\rangle \in \mathcal{H}_1, \qquad \langle \vec{p} | \vec{p}\,' \rangle = \delta(\vec{p} - \vec{p}\,') \, . \tag{25}$$

The states $\{|\vec{p}\rangle\}$ form a basis of $\mathcal{H}_1$ and therefore satisfy the closure relation

$$\int d^3p \, |\vec{p}\rangle \langle \vec{p}| = \mathbf{1} \tag{26}$$

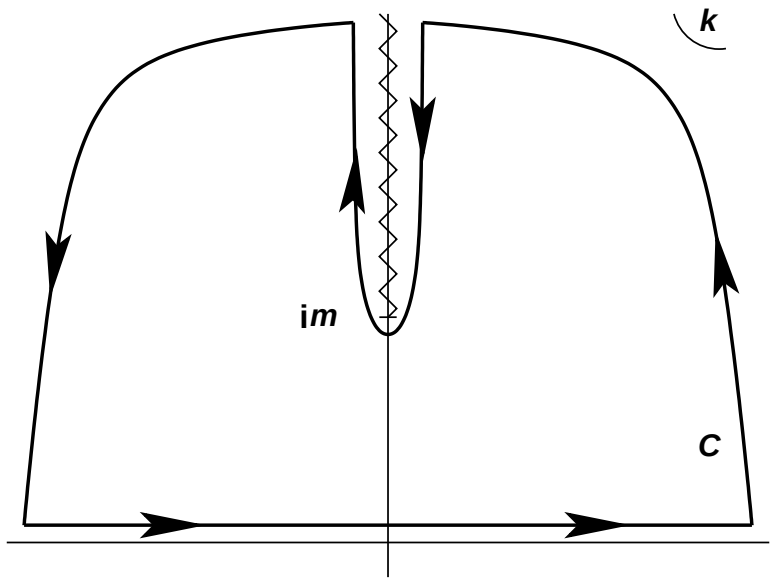


Fig. 5: Complex contour C for the computation of the integral in Eq. (24).

The group of spatial rotations acts unitarily on the states $|\vec{p}\rangle$. This means that for every rotation $R \in \mathrm{SO}(3)$ there is a unitary operator $\mathcal{U}(R)$ such that

$$\mathcal{U}(R)|\vec{p}\rangle = |R\vec{p}\rangle \tag{27}$$

where $R\vec{p}$ represents the action of the rotation on the vector $\vec{k}$, $(R\vec{p})^i = R^i{}_j k^j$. Using a spectral decomposition, the momentum operator $\widehat{P}^i$ can be written as

$$\widehat{P}^i = \int d^3p\, |\vec{p}\rangle\, p^i\, \langle\vec{p}| \tag{28}$$

With the help of Eq. (27) it is straightforward to check that the momentum operator transforms as a vector under rotations:

$$\mathcal{U}(R)^{-1}\, \widehat{P}^i\, \mathcal{U}(R) = \int d^3p\, |R^{-1}\vec{p}\rangle\ p^i\ \langle R^{-1}\vec{p}| = R^i{}_j \widehat{P}^j, \tag{29}$$

where we have used that the integration measure is invariant under $\mathrm{SO}(3)$.

Since, as we argued above, we are forced to deal with multiparticle states, it is convenient to introduce creation-annihilation operators associated with a single-particle state of momentum $\vec{p}$

$$[a(\vec{p}), a^\dagger(\vec{p}\,')] = \delta(\vec{p} - \vec{p}\,'), \qquad [a(\vec{p}), a(\vec{p}\,')] = [a^\dagger(\vec{p}), a^\dagger(\vec{p}\,')] = 0, \tag{30}$$

such that the state $|\vec{p}\rangle$ is created out of the Fock space vacuum $|0\rangle$ (normalized such that $\langle 0|0\rangle = 1$) by the action of a creation operator $a^\dagger(\vec{p})$

$$|\vec{p}\rangle = a^\dagger(\vec{p})|0\rangle, \qquad a(\vec{p})|0\rangle = 0 \quad \forall\vec{p}. \tag{31}$$

Covariance under spatial rotations is all we need if we are interested in a nonrelativistic theory. However in a relativistic quantum field theory we must preserve more that $\mathrm{SO}(3)$, actually we need the expressions to be covariant under the full Poincaré group $\mathrm{ISO}(1,3)$ consisting in spatial rotations, boosts and space-time translations. Therefore, in order to build the Fock space of the theory we need two key ingredients: first an invariant normalization for the states, since we want a normalized state in

one reference frame to be normalized in any other inertial frame. And secondly a relativistic invariant integration measure in momentum space, so the spectral decomposition of operators is covariant under the full Poincaré group.

Let us begin with the invariant measure. Given an invariant function $f(p)$ of the four-momentum p^μ of a particle of mass m with positive energy $p^0 > 0$, there is an integration measure which is invariant under proper Lorentz transformations[2]

$$\int \frac{d^4p}{(2\pi)^4}\,(2\pi)\delta(p^2-m^2)\,\theta(p^0)\,f(p), \tag{32}$$

where $\theta(x)$ represent the Heaviside step function. The integration over p^0 can be easily done using the δ-function identity

$$\delta[f(x)] = \sum_{x_i=\text{zeros of } f} \frac{1}{|f'(x_i)|}\delta(x-x_i), \tag{33}$$

which in our case implies that

$$\delta(p^2-m^2) = \frac{1}{2p^0}\,\delta\left(p^0-\sqrt{\vec{p}^2+m^2}\right) + \frac{1}{2p^0}\,\delta\left(p^0+\sqrt{\vec{p}^2+m^2}\right). \tag{34}$$

The second term in the previous expression correspond to states with negative energy and therefore does not contribute to the integral. We can write then

$$\int \frac{d^4p}{(2\pi)^4}\,(2\pi)\delta(p^2-m^2)\,\theta(p^0)\,f(p) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\sqrt{\vec{p}^2+m^2}}\,f\left(\sqrt{\vec{p}^2+m^2},\vec{p}\right). \tag{35}$$

Hence, the relativistic invariant measure is given by

$$\int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p} \qquad \text{with} \qquad \omega_p \equiv \sqrt{\vec{p}^2+m^2}. \tag{36}$$

Once we have an invariant measure the next step is to find an invariant normalization for the states. We work with a basis $\{|p\rangle\}$ of eigenstates of the four-momentum operator $\widehat{P}^\mu$

$$\widehat{P}^0|p\rangle = \omega_p|p\rangle, \qquad\qquad \widehat{P}^i|p\rangle = p^{\,i}|p\rangle. \tag{37}$$

Since the states $|p\rangle$ are eigenstates of the three-momentum operator we can express them in terms of the non-relativistic states $|\vec{p}\rangle$ that we introduced in Eq. (25)

$$|p\rangle = N(\vec{p})|\vec{p}\rangle \tag{38}$$

with $N(\vec{p})$ a normalization to be determined now. The states $\{|p\rangle\}$ form a complete basis, so they should satisfy the Lorentz invariant closure relation

$$\int \frac{d^4p}{(2\pi)^4}\,(2\pi)\delta(p^2-m^2)\,\theta(p^0)\,|p\rangle\,\langle p| = \mathbf{1} \tag{39}$$

At the same time, this closure relation can be expressed, using Eq. (38), in terms of the nonrelativistic basis of states $\{|\vec{p}\rangle\}$ as

$$\int \frac{d^4p}{(2\pi)^4}\,(2\pi)\delta(p^2-m^2)\,\theta(p^0)\,|p\rangle\,\langle p| = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p}|N(p)|^2\,|\vec{p}\rangle\,\langle\vec{p}|. \tag{40}$$

[2]The factors of 2π are introduced for later convenience.

Using now Eq. (28) for the nonrelativistic states, expression (39) follows provided

$$|N(\vec{p})|^2 = (2\pi)^3\,(2\omega_p). \tag{41}$$

Taking the overall phase in Eq. (38) so that $N(p)$ is real, we define the Lorentz invariant states $|p\rangle$ as

$$|p\rangle = (2\pi)^{\frac{3}{2}}\,\sqrt{2\omega_p}\,|\vec{p}\rangle, \tag{42}$$

and given the normalization of $|\vec{p}\rangle$ we find the normalization of the relativistic states to be

$$\langle p|p'\rangle = (2\pi)^3(2\omega_p)\delta(\vec{p}-\vec{p}\,'). \tag{43}$$

Although not obvious at first sight, the previous normalization is Lorentz invariant. Although it is not difficult to show this in general, here we consider the simpler case of 1+1 dimensions where the two components (p^0, p^1) of the on-shell momentum can be parametrized in terms of a single hyperbolic angle λ as

$$p^0 = m\cosh\lambda, \qquad\qquad p^1 = m\sinh\lambda. \tag{44}$$

Now, the combination $2\omega_p\delta(p^1-p^{1\prime})$ can be written as

$$2\omega_p\delta(p^1-p^{1\prime}) = 2m\cosh\lambda\,\delta(m\sinh\lambda - m\sinh\lambda') = 2\delta(\lambda-\lambda'), \tag{45}$$

where we have made use of the property (33) of the δ-function. Lorentz transformations in $1+1$ dimensions are labelled by a parameter $\xi\in\mathbb{R}$ and act on the momentum by shifting the hyperbolic angle $\lambda\to\lambda+\xi$. However, Eq. (45) is invariant under a common shift of λ and λ', so the whole expression is obviously invariant under Lorentz transformations.

To summarize what we did so far, we have succeed in constructing a Lorentz covariant basis of states for the one-particle Hilbert space $\mathcal{H}_1$. The generators of the Poincaré group act on the states $|p\rangle$ of the basis as

$$\widehat{P}^\mu|p\rangle = p^\mu|p\rangle, \qquad\qquad \mathcal{U}(\Lambda)|p\rangle = |\Lambda^\mu{}_\nu\,p^\nu\rangle \equiv |\Lambda p\rangle \quad \text{with} \quad \Lambda\in \mathrm{SO}(1,3). \tag{46}$$

This is compatible with the Lorentz invariance of the normalization that we have checked above

$$\langle p|p'\rangle = \langle p|\mathcal{U}(\Lambda)^{-1}\mathcal{U}(\Lambda)|p'\rangle = \langle \Lambda p|\Lambda p'\rangle. \tag{47}$$

On $\mathcal{H}_1$ the operator $\widehat{P}^\mu$ admits the following spectral representation

$$\widehat{P}^\mu = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p}|p\rangle\,p^\mu\,\langle p|\,. \tag{48}$$

Using (47) and the fact that the measure is invariant under Lorentz transformation, one can easily show that $\widehat{P}^\mu$ transform covariantly under $\mathrm{SO}(1,3)$

$$\mathcal{U}(\Lambda)^{-1}\widehat{P}^\mu\mathcal{U}(\Lambda) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p}|\Lambda^{-1}p\rangle\,p^\mu\,\langle\Lambda^{-1}p| = \Lambda^\mu{}_\nu\widehat{P}^\nu. \tag{49}$$

A set of covariant creation-annihilation operators can be constructed now in terms of the operators $a(\vec{p})$, $a^\dagger(\vec{p})$ introduced above

$$\alpha(\vec{p}) \equiv (2\pi)^{\frac{3}{2}}\sqrt{2\omega_p}a(\vec{p}), \qquad\qquad \alpha^\dagger(\vec{p}) \equiv (2\pi)^{\frac{3}{2}}\sqrt{2\omega_p}a^\dagger(\vec{p}) \tag{50}$$

with the Lorentz invariant commutation relations

$$[\alpha(\vec{p}),\alpha^\dagger(\vec{p}\,')] \;\;=\;\; (2\pi)^3(2\omega_p)\delta(\vec{p}-\vec{p}\,'),$$

$$[\alpha(\vec{p}),\alpha(\vec{p}\,')] \;=\; [\alpha^\dagger(\vec{p}),\alpha^\dagger(\vec{p}\,')]=0. \tag{51}$$

Particle states are created by acting with any number of creation operators $\alpha(\vec{p})$ on the Poincaré invariant vacuum state $|0\rangle$ satisfying

$$\langle 0|0\rangle = 1, \qquad \widehat{P}^\mu|0\rangle = 0, \qquad \mathcal{U}(\Lambda)|0\rangle = |0\rangle, \quad \forall \Lambda \in \mathrm{SO}(1,3). \tag{52}$$

A general one-particle state $|f\rangle \in \mathcal{H}_1$ can be then written as

$$|f\rangle = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p} f(\vec{p})\alpha^\dagger(\vec{p})|0\rangle, \tag{53}$$

while a n-particle state $|f\rangle \in \mathcal{H}_1^{\otimes n}$ can be expressed as

$$|f\rangle = \int \prod_{i=1}^{n} \frac{d^3p_i}{(2\pi)^3}\frac{1}{2\omega_{p_i}} f(\vec{p}_1,\ldots,\vec{p}_n)\alpha^\dagger(\vec{p}_1)\ldots\alpha^\dagger(\vec{p}_n)|0\rangle. \tag{54}$$

That this states are Lorentz invariant can be checked by noticing that from the definition of the creation-annihilation operators follows the transformation

$$\mathcal{U}(\Lambda)\alpha(\vec{p})\mathcal{U}(\Lambda)^\dagger = \alpha(\Lambda\vec{p}) \tag{55}$$

and the corresponding one for creation operators.

As we have argued above, the very fact that measurements have to be localized implies the necessity of introducing quantum fields. Here we will consider the simplest case of a scalar quantum field $\phi(x)$ satisfying the following properties:

- **Hermiticity.**

$$\phi^\dagger(x) = \phi(x). \tag{56}$$

- **Microcausality.** Since measurements cannot interfere with each other when performed in causally disconnected points of space-time, the commutator of two fields have to vanish outside the relative ligth-cone

$$[\phi(x),\phi(y)] = 0, \qquad (x-y)^2 < 0. \tag{57}$$

- **Translation invariance.**

$$e^{i\widehat{P}\cdot a}\phi(x)e^{-i\widehat{P}\cdot a} = \phi(x-a). \tag{58}$$

- **Lorentz invariance.**

$$\mathcal{U}(\Lambda)^\dagger\phi(x)\mathcal{U}(\Lambda) = \phi(\Lambda^{-1}x). \tag{59}$$

- **Linearity.** To simplify matters we will also assume that $\phi(x)$ is linear in the creation-annihilation operators $\alpha(\vec{p})$, $\alpha^\dagger(\vec{p})$

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p}\left[f(\vec{p},x)\alpha(\vec{p}) + g(\vec{p},x)\alpha^\dagger(\vec{p})\right]. \tag{60}$$

  Since $\phi(x)$ should be hermitian we are forced to take $f(\vec{p},x)^* = g(\vec{p},x)$. Moreover, $\phi(x)$ satisfies the equations of motion of a free scalar field, $(\partial_\mu\partial^\mu + m^2)\phi(x) = 0$, only if $f(\vec{p},x)$ is a complete basis of solutions of the Klein-Gordon equation. These considerations leads to the expansion

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p}\left[e^{-i\omega_p t + i\vec{p}\cdot\vec{x}}\alpha(\vec{p}) + e^{i\omega_p t - i\vec{p}\cdot\vec{x}}\alpha^\dagger(\vec{p})\right]. \tag{61}$$

Given the expansion of the scalar field in terms of the creation-annihilation operators it can be checked that $\phi(x)$ and $\partial_t\phi(x)$ satisfy the equal-time canonical commutation relations

$$[\phi(t,\vec{x}),\partial_t\phi(t,\vec{y})] = i\delta(\vec{x}-\vec{y}) \tag{62}$$

The general commutator $[\phi(x),\phi(y)]$ can be also computed to be

$$[\phi(x),\phi(x')] = i\Delta(x-x'). \tag{63}$$

The function $\Delta(x-y)$ is given by

$$\begin{aligned} i\Delta(x-y) &= -\mathrm{Im}\int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p}e^{-i\omega_p(t-t')+i\vec{p}\cdot(\vec{x}-\vec{x}')} \\ &= \int \frac{d^4p}{(2\pi)^4}(2\pi)\delta(p^2-m^2)\varepsilon(p^0)e^{-ip\cdot(x-x')}, \end{aligned} \tag{64}$$

where $\varepsilon(x)$ is defined as

$$\varepsilon(x) \equiv \theta(x)-\theta(-x) = \left\{ \begin{array}{cc} 1 & x>0 \\ -1 & x<0 \end{array} \right. . \tag{65}$$

Using the last expression in Eq. (64) it is easy to show that $i\Delta(x-x')$ vanishes when x and x' are space-like separated. Indeed, if $(x-x')^2<0$ there is always a reference frame in which both events are simultaneous, and since $i\Delta(x-x')$ is Lorentz invariant we can compute it in this reference frame. In this case $t=t'$ and the exponential in the second line of (64) does not depend on p^0. Therefore, the integration over k^0 gives

$$\begin{aligned} \int_{-\infty}^{\infty} dp^0\varepsilon(p^0)\delta(p^2-m^2) &= \int_{-\infty}^{\infty} dp^0\left[\frac{1}{2\omega_p}\varepsilon(p^0)\delta(p^0-\omega_p)+\frac{1}{2\omega_p}\varepsilon(p^0)\delta(p^0+\omega_p)\right] \\ &= \frac{1}{2\omega_p}-\frac{1}{2\omega_p}=0. \end{aligned} \tag{66}$$

So we have concluded that $i\Delta(x-x')=0$ if $(x-x')^2<0$, as required by microcausality. Notice that the situation is completely different when $(x-x')^2\geq 0$, since in this case the exponential depends on p^0 and the integration over this component of the momentum does not vanish.

3.1 Canonical quantization

So far we have contented ourselves with requiring a number of properties to the quantum scalar field: existence of asymptotic states, locality, microcausality and relativistic invariance. With these only ingredients we have managed to go quite far. The previous can also be obtained using canonical quantization. One starts with a classical free scalar field theory in Hamiltonian formalism and obtains the quantum theory by replacing Poisson brackets by commutators. Since this quantization procedure is based on the use of the canonical formalism, which gives time a privileged rôle, it is important to check at the end of the calculation that the resulting quantum theory is Lorentz invariant. In the following we will briefly overview the canonical quantization of the Klein-Gordon scalar field.

The starting point is the action functional $S[\phi(x)]$ which, in the case of a free real scalar field of mass m is given by

$$S[\phi(x)] \equiv \int d^4x\,\mathcal{L}(\phi,\partial_\mu\phi) = \frac{1}{2}\int d^4x\,\left(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2\right). \tag{67}$$

The equations of motion are obtained, as usual, from the Euler-Lagrange equations

$$\partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right] - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \qquad \Longrightarrow \qquad (\partial_\mu \partial^\mu + m^2)\phi = 0. \tag{68}$$

The momentum canonically conjugated to the field $\phi(x)$ is given by

$$\pi(x) \equiv \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi)} = \frac{\partial \phi}{\partial t}. \tag{69}$$

In the Hamiltonian formalism the physical system is described not in terms of the generalized coordinates and their time derivatives but in terms of the generalized coordinates and their canonically conjugated momenta. This is achieved by a Legendre transformation after which the dynamics of the system is determined by the Hamiltonian function

$$H \equiv \int d^3x \left(\pi \frac{\partial \phi}{\partial t} - \mathcal{L} \right) = \frac{1}{2} \int d^3x \left[\pi^2 + (\vec{\nabla}\phi)^2 + m^2 \right]. \tag{70}$$

The equations of motion can be written in terms of the Poisson rackets. Given two functional $A[\phi, \pi]$, $B[\phi, \pi]$ of the canonical variables

$$A[\phi, \pi] = \int d^3x \mathcal{A}(\phi, \pi), \qquad B[\phi, \pi] = \int d^3x \mathcal{B}(\phi, \pi). \tag{71}$$

Their Poisson bracket is defined by

$$\{A, B\} \equiv \int d^3x \left[\frac{\delta A}{\delta \phi} \frac{\delta B}{\delta \pi} - \frac{\delta A}{\delta \pi} \frac{\delta B}{\delta \phi} \right], \tag{72}$$

where $\frac{\delta}{\delta \phi}$ denotes the functional derivative defined as

$$\frac{\delta A}{\delta \phi} \equiv \frac{\partial \mathcal{A}}{\partial \phi} - \partial_\mu \left[\frac{\partial \mathcal{A}}{\partial(\partial_\mu \phi)} \right] \tag{73}$$

Then, the canonically conjugated fields satisfy the following equal time Poisson brackets

$$\begin{aligned} \{\phi(t, \vec{x}), \phi(t, \vec{x}\,')\} &= \{\pi(t, \vec{x}), \pi(t, \vec{x}\,')\} = 0, \\ \{\phi(t, \vec{x}), \pi(t, \vec{x}\,')\} &= \delta(\vec{x} - \vec{x}\,'). \end{aligned} \tag{74}$$

Canonical quantization proceeds now by replacing classical fields with operators and Poisson brackets with commutators according to the rule

$$i\{\cdot, \cdot\} \longrightarrow [\cdot, \cdot]. \tag{75}$$

In the case of the scalar field, a general solution of the field equations (68) can be obtained by working with the Fourier transform

$$(\partial_\mu \partial^\mu + m^2)\phi(x) = 0 \qquad \Longrightarrow \qquad (-p^2 + m^2)\widetilde{\phi}(p) = 0, \tag{76}$$

whose general solution can be written as[3]

$$\phi(x) = \int \frac{d^4p}{(2\pi)^4} (2\pi)\delta(p^2 - m^2)\theta(p^0) \left[\alpha(p) e^{-ip\cdot x} + \alpha(p)^* e^{ip\cdot x} \right]$$

[3]In momentum space, the general solution to this equation is $\widetilde{\phi}(p) = f(p)\delta(p^2 - m^2)$, with $f(p)$ a completely general function of p^μ. The solution in position space is obtained by inverse Fourier transform.

$$= \int \frac{d^3p}{(2\pi)^3}\frac{1}{2\omega_p}\left[\alpha(\vec{p})e^{-i\omega_p t+\vec{p}\cdot\vec{x}}+\alpha(\vec{p})^* e^{i\omega_p t-\vec{p}\cdot\vec{x}}\right] \tag{77}$$

and we have required $\phi(x)$ to be real. The conjugate momentum is

$$\pi(x) = -\frac{i}{2}\int \frac{d^3p}{(2\pi)^3}\left[\alpha(\vec{p})e^{-i\omega_p t+\vec{p}\cdot\vec{x}}+\alpha(\vec{p})^* e^{i\omega_p t-\vec{p}\cdot\vec{x}}\right]. \tag{78}$$

Now $\phi(x)$ and $\pi(x)$ are promoted to operators by replacing the functions $\alpha(\vec{p})$, $\alpha(\vec{p})^*$ by the corresponding operators

$$\alpha(\vec{p}) \longrightarrow \widehat{\alpha}(\vec{p}), \qquad\qquad \alpha(\vec{p})^* \longrightarrow \widehat{\alpha}^\dagger(\vec{p}). \tag{79}$$

Moreover, demanding $[\phi(t,\vec{x}),\pi(t,\vec{x}\,')] = i\delta(\vec{x}-\vec{x}\,')$ forces the operators $\widehat{\alpha}(\vec{p})$, $\widehat{\alpha}(\vec{p})^\dagger$ to have the commutation relations found in Eq. (51). Therefore they are identified as a set of creation-annihilation operators creating states with well-defined momentum $\vec{p}$ out of the vacuum $|0\rangle$. In the canonical quantization formalism the concept of particle appears as a result of the quantization of a classical field.

Knowing the expressions of $\widehat{\phi}$ and $\widehat{\pi}$ in terms of the creation-annihilation operators we can proceed to evaluate the Hamiltonian operator. After a simple calculation one arrives to the expression

$$\widehat{H} = \int d^3p\,\left[\omega_p \widehat{\alpha}^\dagger(\vec{p})\widehat{\alpha}(\vec{p}) + \frac{1}{2}\omega_p\,\delta(\vec{0})\right]. \tag{80}$$

The first term has a simple physical interpretation since $\widehat{\alpha}^\dagger(\vec{p})\widehat{\alpha}(\vec{p})$ is the number operator of particles with momentum $\vec{p}$. The second divergent term can be eliminated if we defined the normal-ordered Hamiltonian $:\widehat{H}:$ with the vacuum energy subtracted

$$:\widehat{H}:\equiv \widehat{H} - \langle 0|\widehat{H}|0\rangle = \int d^3p\,\omega_p\,\widehat{\alpha}^\dagger(\vec{p})\,\widehat{\alpha}(\vec{p}) \tag{81}$$

It is interesting to try to make sense of the divergent term in Eq. (80). This term have two sources of divergence. One is associated with the delta function evaluated at zero coming from the fact that we are working in a infinite volume. It can be regularized for large but finite volume by replacing $\delta(\vec{0}) \sim V$. Hence, it is of infrared origin. The second one comes from the integration of ω_p at large values of the momentum and it is then an ultraviolet divergence. The infrared divergence can be regularized by considering the scalar field to be living in a box of finite volume V. In this case the vacuum energy is

$$E_{\text{vac}} \equiv \langle 0|\widehat{H}|0\rangle = \sum_{\vec{p}}\frac{1}{2}\omega_p. \tag{82}$$

Written in this way the interpretation of the vacuum energy is straightforward. A free scalar quantum field can be seen as a infinite collection of harmonic oscillators per unit volume, each one labelled by $\vec{p}$. Even if those oscillators are not excited, they contribute to the vacuum energy with their zero-point energy, given by $\frac{1}{2}\omega_p$. This vacuum contribution to the energy add up to infinity even if we work at finite volume, since even then there are modes with arbitrary high momentum contributing to the sum, $p_i = \frac{n_i\pi}{L_i}$, with L_i the sides of the box of volume V and n_i an integer. Hence, this divergence is of ultraviolet origin.

Our discussion leads us to the conclusion that the vacuum in quantum field theory is radically different from the classical idea of the vacuum as "empty space". Indeed, we have seen that a quantum field can be regarded as a set of an infinite number of harmonic oscillators and that the ground state of the system is obtained when *all* oscillators are in their respective ground states. This being so, we know from elementary quantum mechanics that a harmonic oscillator in its ground state is not "at rest", but

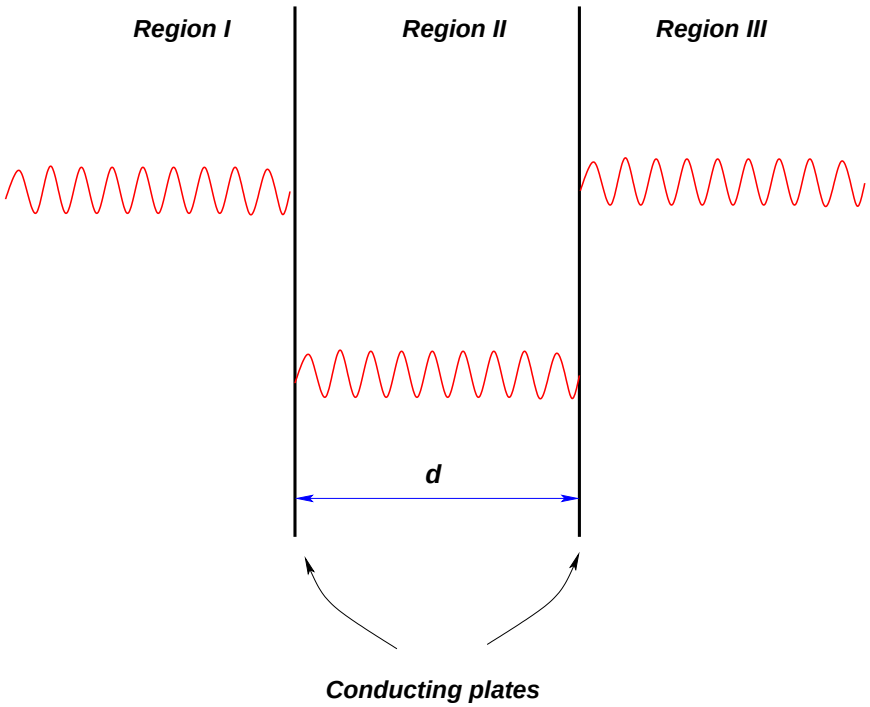


Fig. 6: Illustration of the Casimir effect. In regions I and II the spetrum of modes of the momentum $p_\perp$ is continuous, while in the space between the plates (region II) it is quantized in units of $\frac{\pi}{d}$.

fluctuate with an energy given by its zero-point energy. When translated to quantum field theory, this means that the vacuum can be picture as a medium where virtual particles are continuously created and annihilated. As we will see, this nontrivial character of the vacuum has physical consequences ranging from the Casimir effect (see below) to the screening or antiscreening of charges in gauge theories (see Section 8.2).

3.2 The Casimir effect

The presence of a vacuum energy is not characteristic of the scalar field. It is also present in other cases, in particular in quantum electrodynamics. Although one might be tempted to discarding this infinite contribution to the energy of the vacuum as unphysical, it has observable consequences. In 1948 Hendrik Casimir pointed out [16] that although a formally divergent vacuum energy would not be observable, any variation in this energy would be (see [17] for comprehensive reviews).

To show this he devised the following experiment. Consider a couple of infinite, perfectly conducting plates placed parallel to each other at a distance d (see Fig. 6). Because the conducting plates fix the boundary condition of the vacuum modes of the electromagnetic field these are discrete in between the plates (region II), while outside there is a continuous spectrum of modes (regions I and III). In order to calculate the force between the plates we can take the vacuum energy of the electromagnetic field as given by the contribution of two scalar fields corresponding to the two polarizations of the photon. Therefore we can use the formulas derived above.

A naive calculation of the vacuum energy in this system gives a divergent result. This infinity can be removed, however, by substracting the vacuum energy corresponding to the situation where the plates are removed

$$E(d)_{\rm reg} = E(d)_{\rm vac} - E(\infty)_{\rm vac} \tag{83}$$

This substraction cancels the contribution of the modes outside the plates. Because of the boundary

conditions imposed by the plates the momentum of the modes perpendicular to the plates are quantized according to $p_\perp = \frac{n\pi}{d}$, with n a non-negative integer. If we consider that the size of the plates is much larger than their separation d we can take the momenta parallel to the plates $\vec{p}_\parallel$ as continuous. For $n > 0$ we have two polarizations for each vacuum mode of the electromagnetic field, each contributing like $\frac{1}{2}\sqrt{\vec{p}_\parallel^{\,2} + p_\perp^2}$ to the vacuum energy. On the other hand, when $p_\perp = 0$ the corresponding modes of the field are effectively (2+1)-dimensional and therefore there is only one polarization. Keeping this in mind, we can write

$$\begin{aligned} E(d)_{\rm reg} &= S\int \frac{d^2p_\parallel}{(2\pi)^2}\frac{1}{2}|\vec{p}_\parallel| + 2S\int \frac{d^2p_\parallel}{(2\pi)^2}\sum_{n=1}^{\infty}\frac{1}{2}\sqrt{\vec{p}_\parallel^{\,2} + \left(\frac{n\pi}{d}\right)^2} \\ &- 2Sd\int\frac{d^3p}{(2\pi)^3}\frac{1}{2}|\vec{p}| \end{aligned} \tag{84}$$

where S is the area of the plates. The factors of 2 take into account the two propagating degrees of freedom of the electromagnetic field, as discussed above. In order to ensure the convergence of integrals and infinite sums we can introduce an exponential damping factor[4]

$$\begin{aligned} E(d)_{\rm reg} &= \frac{1}{2}S\int\frac{d^2p_\perp}{(2\pi)^2}e^{-\frac{1}{\Lambda}|\vec{p}_\parallel|}|\vec{p}_\parallel| + S\sum_{n=1}^{\infty}\int\frac{d^2p_\parallel}{(2\pi)^2}e^{-\frac{1}{\Lambda}\sqrt{\vec{p}_\parallel^{\,2}+\left(\frac{n\pi}{d}\right)^2}}\sqrt{\vec{p}_\parallel^{\,2} + \left(\frac{n\pi}{d}\right)^2} \\ &- Sd\int_{-\infty}^{\infty}\frac{dp_\perp}{2\pi}\int\frac{d^2p_\parallel}{(2\pi)^2}e^{-\frac{1}{\Lambda}\sqrt{\vec{p}_\parallel^{\,2}+p_\perp^2}}\sqrt{\vec{p}_\parallel^{\,2}+p_\perp^2} \end{aligned} \tag{85}$$

where Λ is an ultraviolet cutoff. It is now straightforward to see that if we define the function

$$F(x) = \frac{1}{2\pi}\int_0^\infty y\,dy\,e^{-\frac{1}{\Lambda}\sqrt{y^2+\left(\frac{x\pi}{d}\right)^2}}\sqrt{y^2+\left(\frac{x\pi}{d}\right)^2} = \frac{1}{4\pi}\int_{\left(\frac{x\pi}{d}\right)^2}^\infty dz\,e^{-\frac{\sqrt{z}}{\Lambda}}\sqrt{z} \tag{86}$$

the regularized vacuum energy can be written as

$$E(d)_{\rm reg} = S\left[\frac{1}{2}F(0) + \sum_{n=1}^{\infty}F(n) - \int_0^\infty dx\,F(x)\right] \tag{87}$$

This expression can be evaluated using the Euler-MacLaurin formula [19]

$$\begin{aligned} \sum_{n=1}^{\infty}F(n) - \int_0^\infty dx\,F(x) &= -\frac{1}{2}\left[F(0)+F(\infty)\right] + \frac{1}{12}\left[F'(\infty)-F'(0)\right] \\ &- \frac{1}{720}\left[F'''(\infty)-F'''(0)\right] + \dots \end{aligned} \tag{88}$$

Since for our function $F(\infty) = F'(\infty) = F'''(\infty) = 0$ and $F'(0) = 0$, the value of $E(d)_{\rm reg}$ is determined by $F'''(0)$. Computing this term and removing the ultraviolet cutoff, $\Lambda \to \infty$ we find the result

$$E(d)_{\rm reg} = \frac{S}{720}F'''(0) = -\frac{\pi^2 S}{720 d^3}. \tag{89}$$

Then, the force per unit area between the plates is given by

$$P_{\rm Casimir} = -\frac{\pi^2}{240}\frac{1}{d^4}. \tag{90}$$

The minus sign shows that the force between the plates is attractive. This is the so-called Casimir effect. It was experimentally measured in 1958 by Sparnaay [18] and since then the Casimir effect has been checked with better and better precission in a variety of situations [17].

[4] Actually, one could introduce any cutoff function $f(p_\perp^2 + p_\parallel^2)$ going to zero fast enough as $p_\perp$, $p_\parallel \to \infty$. The result is independent of the particular function used in the calculation.

4 Theories and Lagrangians

Up to this point we have used a scalar field to illustrate our discussion of the quantization procedure. However, nature is richer than that and it is necessary to consider other fields with more complicated behavior under Lorentz transformations. Before considering other fields we pause and study the properties of the Lorentz group.

4.1 Representations of the Lorentz group

In four dimensions the Lorentz group has six generators. Three of them correspond to the generators of the group of rotations in three dimensions SO(3). In terms of the generators J_i of the group a finite rotation of angle φ with respect to an axis determined by a unitary vector $\vec{e}$ can be written as

$$R(\vec{e},\varphi) = e^{-i\varphi\,\vec{e}\cdot\vec{J}}, \qquad \vec{J} = \begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix}. \tag{91}$$

The other three generators of the Lorentz group are associated with boosts M_i along the three spatial directions. A boost with rapidity λ along a direction $\vec{u}$ is given by

$$B(\vec{u},\lambda) = e^{-i\lambda\,\vec{u}\cdot\vec{M}}, \qquad \vec{M} = \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix}. \tag{92}$$

These six generators satisfy the algebra

$$\begin{aligned} [J_i, J_j] &= i\epsilon_{ijk}J_k, \\ [J_i, M_k] &= i\epsilon_{ijk}M_k, \\ [M_i, M_j] &= -i\epsilon_{ijk}J_k, \end{aligned} \tag{93}$$

The first line corresponds to the commutation relations of SO(3), while the second one implies that the generators of the boosts transform like a vector under rotations.

At first sight, to find representations of the algebra (93) might seem difficult. The problem is greatly simplified if we consider the following combination of the generators

$$J_k^{\pm} = \frac{1}{2}(J_k \pm iM_k). \tag{94}$$

Using (93) it is easy to prove that the new generators $J_k^{\pm}$ satisfy the algebra

$$\begin{aligned} [J_i^{\pm}, J_j^{\pm}] &= i\epsilon_{ijk}J_k^{\pm}, \\ [J_i^{+}, J_j^{-}] &= 0. \end{aligned} \tag{95}$$

Then the Lorentz algebra (93) is actually equivalent to two copies of the algebra of $\mathrm{SU}(2) \approx \mathrm{SO}(3)$. Therefore the irreducible representations of the Lorentz group can be obtained from the well-known representations of SU(2). Since the latter ones are labelled by the spin $\mathbf{s} = k + \frac{1}{2}, k$ (with $k \in \mathbb{N}$), any representation of the Lorentz algebra can be identified by specifying $(\mathbf{s}_+, \mathbf{s}_-)$, the spins of the representations of the two copies of SU(2) that made up the algebra (93).

To get familiar with this way of labelling the representations of the Lorentz group we study some particular examples. Let us start with the simplest one $(\mathbf{s}_+, \mathbf{s}_-) = (\mathbf{0}, \mathbf{0})$. This state is a singlet under $J_i^{\pm}$ and therefore also under rotations and boosts. Therefore we have a scalar.

The next interesting cases are $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$ and $(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$. They correspond respectively to a right-handed and a left-handed Weyl spinor. Their properties will be studied in more detail below. In the case of

Representation	Type of field
$(\mathbf{0},\mathbf{0})$	Scalar
$(\frac{\mathbf{1}}{\mathbf{2}},\mathbf{0})$	Right-handed spinor
$(\mathbf{0},\frac{\mathbf{1}}{\mathbf{2}})$	Left-handed spinor
$(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}})$	Vector
$(\mathbf{1},\mathbf{0})$	Selfdual antisymmetric 2-tensor
$(\mathbf{0},\mathbf{1})$	Anti-selfdual antisymmetric 2-tensor

Table 1: Representations of the Lorentz group

$(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}})$, since from Eq. (94) we see that $J_i = J_i^+ + J_i^-$ the rules of addition of angular momentum tell us that there are two states, one of them transforming as a vector and another one as a scalar under three-dimensional rotations. Actually, a more detailed analysis shows that the singlet state corresponds to the time component of a vector and the states combine to form a vector under the Lorentz group.

There are also more "exotic" representations. For example we can consider the $(\mathbf{1},\mathbf{0})$ and $(\mathbf{0},\mathbf{1})$ representations corresponding respectively to a selfdual and an anti-selfdual rank-two antisymmetric tensor. In Table 1 we summarize the previous discussion.

To conclude our discussion of the representations of the Lorentz group we notice that under a parity transformation the generators of SO(1,3) transform as

$$P : J_i \longrightarrow J_i, \qquad P : M_i \longrightarrow -M_i \tag{96}$$

this means that $P : J_i^{\pm} \longrightarrow J_i^{\mp}$ and therefore a representation $(\mathbf{s_1},\mathbf{s_2})$ is transformed into $(\mathbf{s_2},\mathbf{s_1})$. This means that, for example, a vector $(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}})$ is invariant under parity, whereas a left-handed Weyl spinor $(\frac{\mathbf{1}}{\mathbf{2}},\mathbf{0})$ transforms into a right-handed one $(\mathbf{0},\frac{\mathbf{1}}{\mathbf{2}})$ and vice versa.

4.2 Spinors

Weyl spinors. Let us go back to the two spinor representations of the Lorentz group, namely $(\frac{\mathbf{1}}{\mathbf{2}},\mathbf{0})$ and $(\mathbf{0},\frac{\mathbf{1}}{\mathbf{2}})$. These representations can be explicitly constructed using the Pauli matrices as

$$\begin{aligned} J_i^+ &= \frac{1}{2}\sigma^i, \qquad J_i^- = 0 \qquad \text{for} \quad (\tfrac{\mathbf{1}}{\mathbf{2}},\mathbf{0}), \\ J_i^+ &= 0, \qquad J_i^- = \frac{1}{2}\sigma^i \qquad \text{for} \quad (\mathbf{0},\tfrac{\mathbf{1}}{\mathbf{2}}). \end{aligned} \tag{97}$$

We denote by $u_\pm$ a complex two-component object that transforms in the representation $\mathbf{s}_\pm = \frac{1}{2}$ of $J_\pm^i$. If we define $\sigma_\pm^\mu = (\mathbf{1},\pm\sigma^i)$ we can construct the following vector quantities

$$u_+^\dagger \sigma_+^\mu u_+, \qquad u_-^\dagger \sigma_-^\mu u_-. \tag{98}$$

Notice that since $(J_i^\pm)^\dagger = J_i^\mp$ the hermitian conjugated fields $u_\pm^\dagger$ are in the $(\mathbf{0},\frac{\mathbf{1}}{\mathbf{2}})$ and $(\frac{\mathbf{1}}{\mathbf{2}},\mathbf{0})$ respectively.

To construct a free Lagrangian for the fields $u_\pm$ we have to look for quadratic combinations of the fields that are Lorentz scalars. If we also demand invariance under global phase rotations

$$u_\pm \longrightarrow e^{i\theta} u_\pm \tag{99}$$

we are left with just one possibility up to a sign

$$\mathcal{L}^{\pm}_{\text{Weyl}} = iu^{\dagger}_{\pm}\left(\partial_t \pm \vec{\sigma}\cdot\vec{\nabla}\right)u_{\pm} = iu^{\dagger}_{\pm}\sigma^{\mu}_{\pm}\partial_{\mu}u_{\pm}. \tag{100}$$

This is the Weyl Lagrangian. In order to grasp the physical meaning of the spinors $u_{\pm}$ we write the equations of motion

$$\left(\partial_0 \pm \vec{\sigma}\cdot\vec{\nabla}\right)u_{\pm} = 0. \tag{101}$$

Multiplying this equation on the left by $\left(\partial_0 \mp \vec{\sigma}\cdot\vec{\nabla}\right)$ and applying the algebraic properties of the Pauli matrices we conclude that $u_{\pm}$ satisfies the massless Klein-Gordon equation

$$\partial_{\mu}\partial^{\mu}\, u_{\pm} = 0, \tag{102}$$

whose solutions are:

$$u_{\pm}(x) = u_{\pm}(k)e^{-ik\cdot x}, \qquad \text{with} \quad k^0 = |\vec{k}|. \tag{103}$$

Plugging these solutions back into the equations of motion (101) we find

$$\left(|\vec{k}| \mp \vec{k}\cdot\vec{\sigma}\right)u_{\pm} = 0, \tag{104}$$

which implies

$$\begin{aligned} u_+ : &\qquad \frac{\vec{\sigma}\cdot\vec{k}}{|\vec{k}|} = 1, \\ u_- : &\qquad \frac{\vec{\sigma}\cdot\vec{k}}{|\vec{k}|} = -1. \end{aligned} \tag{105}$$

Since the spin operator is defined as $\vec{s} = \frac{1}{2}\vec{\sigma}$, the previous expressions give the chirality of the states with wave function $u_{\pm}$, i.e. the projection of spin along the momentum of the particle. Therefore we conclude that u_+ is a Weyl spinor of positive helicity $\lambda = \frac{1}{2}$, while u_- has negative helicity $\lambda = -\frac{1}{2}$. This agrees with our assertion that the representation $(\mathbf{\frac{1}{2}}, \mathbf{0})$ corresponds to a right-handed Weyl fermion (positive chirality) whereas $(\mathbf{0}, \mathbf{\frac{1}{2}})$ is a left-handed Weyl fermion (negative chirality). For example, in the standard model neutrinos are left-handed Weyl spinors and therefore transform in the representation $(\mathbf{0}, \mathbf{\frac{1}{2}})$ of the Lorentz group.

Nevertheless, it is possible that we were too restrictive in constructing the Weyl Lagrangian (100). There we constructed the invariants from the vector bilinears (98) corresponding to the product representations

$$(\mathbf{\tfrac{1}{2}}, \mathbf{\tfrac{1}{2}}) = (\mathbf{\tfrac{1}{2}}, \mathbf{0}) \otimes (\mathbf{0}, \mathbf{\tfrac{1}{2}}) \quad \text{and} \quad (\tfrac{1}{2}, \tfrac{1}{2}) = (\mathbf{0}, \mathbf{\tfrac{1}{2}}) \otimes (\mathbf{\tfrac{1}{2}}, \mathbf{0}). \tag{106}$$

In particular our insistence in demanding the Lagrangian to be invariant under the global symmetry $u_{\pm} \to e^{i\theta}u_{\pm}$ rules out the scalar term that appears in the product representations

$$(\mathbf{\tfrac{1}{2}}, \mathbf{0}) \otimes (\mathbf{\tfrac{1}{2}}, \mathbf{0}) = (\mathbf{1}, \mathbf{0}) \oplus (\mathbf{0}, \mathbf{0}), \qquad (\mathbf{0}, \mathbf{\tfrac{1}{2}}) \otimes (\mathbf{0}, \mathbf{\tfrac{1}{2}}) = (\mathbf{0}, \mathbf{1}) \oplus (\mathbf{0}, \mathbf{0}). \tag{107}$$

The singlet representations corresponds to the antisymmetric combinations

$$\epsilon_{ab}u^{a}_{\pm}u^{b}_{\pm}, \tag{108}$$

where ϵ_{ab} is the antisymmetric symbol $\epsilon_{12} = -\epsilon_{21} = 1$.

At first sight it might seem that the term (108) vanishes identically because of the antisymmetry of the ϵ-symbol. However we should keep in mind that the spin-statistic theorem (more on this later) demands that fields with half-integer spin have to satisfy the Fermi-Dirac statistics and therefore satisfy anticommutation relations, whereas fields of integer spin follow the statistic of Bose-Einstein and, as a consequence, quantization replaces Poisson brackets by commutators. This implies that the components of the Weyl fermions $u_\pm$ are anticommuting Grassmann fields

$$u^a_\pm u^b_\pm + u^b_\pm u^a_\pm = 0. \tag{109}$$

It is important to realize that, strictly speaking, fermions (i.e., objects that satisfy the Fermi-Dirac statistics) do not exist classically. The reason is that they satisfy the Pauli exclusion principle and therefore each quantum state can be occupied, at most, by one fermion. Therefore the naïve definition of the classical limit as a limit of large occupation numbers cannot be applied. Fermion field do not really make sense classically.

Since the combination (108) does not vanish and we can construct a new Lagrangian

$$\mathcal{L}^\pm_{\rm Weyl} = iu^\dagger_\pm \sigma^\mu_\pm \partial_\mu u_\pm - \frac{m}{2}\epsilon_{ab} u^a_\pm u^b_\pm + \text{h.c.} \tag{110}$$

This mass term, called of Majorana type, is allowed if we do not worry about breaking the global U(1) symmetry $u_\pm \to e^{i\theta} u_\pm$. This is not the case, for example, of charged chiral fermions, since the Majorana mass violates the conservation of electric charge or any other gauge U(1) charge. In the standard model, however, there is no such a problem if we introduce Majorana masses for right-handed neutrinos, since they are singlet under all standard model gauge groups. Such a term will break, however, the global U(1) lepton number charge because the operator $\epsilon_{ab}\nu^a_R \nu^b_R$ changes the lepton number by two units

Dirac spinors. We have seen that parity interchanges the representations $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$ and $(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$, i.e. it changes right-handed with left-handed fermions

$$P : u_\pm \longrightarrow u_\mp. \tag{111}$$

An obvious way to build a parity invariant theory is to introduce a pair or Weyl fermions u_+ and u_+. Actually, these two fields can be combined in a single four-component spinor

$$\psi = \begin{pmatrix} u_+ \\ u_- \end{pmatrix} \tag{112}$$

transforming in the reducible representation $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0}) \oplus (\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$.

Since now we have both u_+ and u_- simultaneously at our disposal the equations of motion for $u_\pm$, $i\sigma^\mu_\pm \partial_\mu u_\pm = 0$ can be modified, while keeping them linear, to

$$\left.\begin{aligned} i\sigma^\mu_+ \partial_\mu u_+ &= m u_- \\ i\sigma^\mu_- \partial_\mu u_- &= m u_+ \end{aligned}\right\} \quad \Longrightarrow \quad i \begin{pmatrix} \sigma^\mu_+ & 0 \\ 0 & \sigma^\mu_- \end{pmatrix} \partial_\mu \psi = m \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \psi. \tag{113}$$

These equations of motion can be derived from the Lagrangian density

$$\mathcal{L}_{\rm Dirac} = i\psi^\dagger \begin{pmatrix} \sigma^\mu_+ & 0 \\ 0 & \sigma^\mu_- \end{pmatrix} \partial_\mu \psi - m\psi^\dagger \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \psi. \tag{114}$$

To simplify the notation it is useful to define the Dirac γ-matrices as

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu_- \\ \sigma^\mu_+ & 0 \end{pmatrix} \tag{115}$$

and the Dirac conjugate spinor $\overline{\psi}$

$$\overline{\psi} \equiv \psi^\dagger \gamma^0 = \psi^\dagger \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}. \tag{116}$$

Now the Lagrangian (114) can be written in the more compact form

$$\mathcal{L}_{\rm Dirac} = \overline{\psi}\,(i\gamma^\mu \partial_\mu - m)\,\psi. \tag{117}$$

The associated equations of motion give the Dirac equation (11) with the identifications

$$\gamma^0 = \beta, \qquad \gamma^i = i\alpha^i. \tag{118}$$

In addition, the γ-matrices defined in (115) satisfy the Clifford algebra

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}. \tag{119}$$

In D dimensions this algebra admits representations of dimension $2^{[\frac{D}{2}]}$. When D is even the Dirac fermions ψ transform in a reducible representation of the Lorentz group. In the case of interest, $D = 4$ this is easy to prove by defining the matrix

$$\gamma^5 = -i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}. \tag{120}$$

We see that γ^5 anticommutes with all other γ-matrices. This implies that

$$[\gamma^5, \sigma^{\mu\nu}] = 0, \qquad \text{with} \qquad \sigma^{\mu\nu} = -\frac{i}{4}[\gamma^\mu, \gamma^\nu]. \tag{121}$$

Because of Schur's lemma (see Appendix) this implies that the representation of the Lorentz group provided by $\sigma^{\mu\nu}$ is reducible into subspaces spanned by the eigenvectors of γ^5 with the same eigenvalue. If we define the projectors $P_\pm = \frac{1}{2}(1 \pm \gamma^5)$ these subspaces correspond to

$$P_+\psi = \begin{pmatrix} u_+ \\ 0 \end{pmatrix}, \qquad P_-\psi = \begin{pmatrix} 0 \\ u_- \end{pmatrix}, \tag{122}$$

which are precisely the Weyl spinors introduced before.

Our next task is to quantize the Dirac Lagrangian. This will be done along the lines used for the Klein-Gordon field, starting with a general solution to the Dirac equation and introducing the corresponding set of creation-annihilation operators. Therefore we start by looking for a complete basis of solutions to the Dirac equation. In the case of the scalar field the elements of the basis were labelled by their four-momentum k^μ. Now, however, we have more degrees of freedom since we are dealing with a spinor which means that we have to add extra labels. Looking back at Eq. (105) we can define the helicity operator for a Dirac spinor as

$$\lambda = \frac{1}{2}\vec{\sigma}\cdot\frac{\vec{k}}{|\vec{k}|}\begin{pmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{pmatrix}. \tag{123}$$

Hence, each element of the basis of functions is labelled by its four-momentum k^μ and the corresponding eigenvalue s of the helicity operator. For positive energy solutions we then propose the ansatz

$$u(k,s)e^{-ik\cdot x}, \qquad s = \pm\frac{1}{2}, \tag{124}$$

where $u_\alpha(k,s)$ ($\alpha = 1,\ldots,4$) is a four-component spinor. Substituting in the Dirac equation we obtain

$$(\not{k} - m)u(k,s) = 0. \tag{125}$$

In the same way, for negative energy solutions we have

$$v(k,s)e^{ik\cdot x}, \qquad s=\pm\frac{1}{2}, \tag{126}$$

where $v(k,s)$ has to satisfy

$$(\not{k}+m)v(k,s)=0. \tag{127}$$

Multiplying Eqs. (125) and (127) on the left respectively by $(\not{k}\mp m)$ we find that the momentum is on the mass shell, $k^2=m^2$. Because of this, the wave function for both positive- and negative-energy solutions can be labeled as well using the three-momentum $\vec{k}$ of the particle, $u(\vec{k},s)$, $v(\vec{k},s)$.

A detailed analysis shows that the functions $u(\vec{k},s)$, $v(\vec{k},s)$ satisfy the properties

$$\begin{aligned} \overline{u}(\vec{k},s)u(\vec{k},s)&=2m, & \overline{v}(\vec{k},s)v(\vec{k},s)&=-2m,\\ \overline{u}(\vec{k},s)\gamma^\mu u(\vec{k},s)&=2k^\mu, & \overline{v}(\vec{k},s)\gamma^\mu v(\vec{k},s)&=2k^\mu,\\ \sum_{s=\pm\frac{1}{2}} u_\alpha(\vec{k},s)\overline{u}_\beta(\vec{k},s)&=(\not{k}+m)_{\alpha\beta}, & \sum_{s=\pm\frac{1}{2}} v_\alpha(\vec{k},s)\overline{v}_\beta(\vec{k},s)&=(\not{k}-m)_{\alpha\beta}, \end{aligned} \tag{128}$$

with $k^0=\omega_k=\sqrt{\vec{k}^{\,2}+m^2}$. Then, a general solution to the Dirac equation including creation and annihilation operators can be written as:

$$\widehat{\psi}(t,\vec{x})=\int\frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_k}\sum_{s=\pm\frac{1}{2}}\left[u(\vec{k},s)\,\widehat{b}(\vec{k},s)e^{-i\omega_k t+i\vec{k}\cdot\vec{x}}+v(\vec{k},s)\,\widehat{d}^\dagger(\vec{k},s)e^{i\omega_k t-i\vec{k}\cdot\vec{x}}\right]. \tag{129}$$

The operators $\widehat{b}^\dagger(\vec{k},s)$, $\widehat{b}(\vec{k})$ respectively create and annihilate a spin-$\frac{1}{2}$ particle (for example, an electron) out of the vacuum with momentum $\vec{k}$ and helicity s. Because we are dealing with half-integer spin fields, the spin-statistics theorem forces canonical anticommutation relations for $\widehat{\psi}$ which means that the creation-annihilation operators satisfy the algebra[5]

$$\begin{aligned} \{b(\vec{k},s),b^\dagger(\vec{k}\,',s')\}&=\delta(\vec{k}-\vec{k}\,')\delta_{ss'},\\ \{b(\vec{k},s),b(\vec{k}\,',s')\}&=\{b^\dagger(\vec{k},s),b^\dagger(\vec{k}\,',s')\}=0. \end{aligned} \tag{130}$$

In the case of $d(\vec{k},s)$, $d^\dagger(\vec{k},s)$ we have a set of creation-annihilation operators for the corresponding antiparticles (for example positrons). This is clear if we notice that $d^\dagger(\vec{k},s)$ can be seen as the annihilation operator of a negative energy state of the Dirac equation with wave function $v_\alpha(\vec{k},s)$. As we saw, in the Dirac sea picture this corresponds to the creation of an antiparticle out of the vacuum (see Fig. 2). The creation-annihilation operators for antiparticles also satisfy the fermionic algebra

$$\begin{aligned} \{d(\vec{k},s),d^\dagger(\vec{k}\,',s')\}&=\delta(\vec{k}-\vec{k}\,')\delta_{ss'},\\ \{d(\vec{k},s),d(\vec{k}\,',s')\}&=\{d^\dagger(\vec{k},s),d^\dagger(\vec{k}\,',s')\}=0. \end{aligned} \tag{131}$$

All other anticommutators between $b(\vec{k},s)$, $b^\dagger(\vec{k},s)$ and $d(\vec{k},s)$, $d^\dagger(\vec{k},s)$ vanish.

The Hamiltonian operator for the Dirac field is

$$\widehat{H}=\frac{1}{2}\sum_{s=\pm\frac{1}{2}}\int\frac{d^3k}{(2\pi)^3}\left[b^\dagger(\vec{k},s)b(\vec{k},s)-d(\vec{k},s)d^\dagger(\vec{k},s)\right]. \tag{132}$$

At this point we realize again of the necessity of quantizing the theory using anticommutators instead of commutators. Had we use canonical commutation relations, the second term inside the integral in

[5] To simplify notation, and since there is no risk of confusion, we drop from now on the hat to indicate operators.

(132) would give the number operator $d^\dagger(\vec{k},s)d(\vec{k},s)$ with a minus sign in front. As a consequence the Hamiltonian would be unbounded from below and we would be facing again the instability of the theory already noticed in the context of relativistic quantum mechanics. However, because of the *anticommutation* relations (131), the Hamiltonian (132) takes the form

$$\widehat{H} = \sum_{s=\pm\frac{1}{2}} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left[\omega_k b^\dagger(\vec{k},s)b(\vec{k},s) + \omega_k d^\dagger(\vec{k},s)d(\vec{k},s)\right] - 2\int d^3k\, \omega_k \delta(\vec{0}). \tag{133}$$

As with the scalar field, we find a divergent vacuum energy contribution due to the zero-point energy of the infinite number of harmonic oscillators. Unlike the Klein-Gordon field, the vacuum energy is negative. In section 9.2 we will see that in certain type of theories called supersymmetric, where the number of bosonic and fermionic degrees of freedom is the same, there is a cancellation of the vacuum energy. The divergent contribution can be removed by the normal order prescription

$$:\widehat{H}:= \sum_{s=\pm\frac{1}{2}} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left[\omega_k b^\dagger(\vec{k},s)b(\vec{k},s) + \omega_k d^\dagger(\vec{k},s)d(\vec{k},s)\right]. \tag{134}$$

Finally, let us mention that using the Dirac equation it is easy to prove that there is a conserved four-current given by

$$j^\mu = \overline{\psi}\gamma^\mu\psi, \qquad \partial_\mu j^\mu = 0. \tag{135}$$

As we will explain further in sec. 6 this current is associated to the invariance of the Dirac Lagrangian under the global phase shift $\psi \to e^{i\theta}\psi$. In electrodynamics the associated conserved charge

$$Q = e\int d^3x\, j^0 \tag{136}$$

is identified with the electric charge.

4.3 Gauge fields

In classical electrodynamics the basic quantities are the electric and magnetic fields $\vec{E}$, $\vec{B}$. These can be expressed in terms of the scalar and vector potential $(\varphi, \vec{A})$

$$\begin{aligned} \vec{E} &= -\vec{\nabla}\varphi - \frac{\partial\vec{A}}{\partial t}, \\ \vec{B} &= \vec{\nabla}\times\vec{A}. \end{aligned} \tag{137}$$

From these equations it follows that there is an ambiguity in the definition of the potentials given by the gauge transformations

$$\varphi(t,\vec{x}) \to \varphi(t,\vec{x}) + \frac{\partial}{\partial t}\epsilon(t,\vec{x}), \qquad \vec{A}(t,\vec{x}) \to \vec{A}(t,\vec{x}) - \vec{\nabla}\epsilon(t,\vec{x}). \tag{138}$$

Classically $(\varphi, \vec{A})$ are seen as only a convenient way to solve the Maxwell equations, but without physical relevance.

The equations of electrodynamics can be recast in a manifestly Lorentz invariant form using the four-vector gauge potential $A^\mu = (\varphi, \vec{A})$ and the antisymmetric rank-two tensor: $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Maxwell's equations become

$$\begin{aligned} \partial_\mu F^{\mu\nu} &= j^\mu, \\ \epsilon^{\mu\nu\sigma\eta}\partial_\nu F_{\sigma\eta} &= 0, \end{aligned} \tag{139}$$

where the four-current $j^\mu = (\rho, \vec{j})$ contains the charge density and the electric current. The field strength tensor $F_{\mu\nu}$ and the Maxwell equations are invariant under gauge transformations (138), which in covariant form read

$$A_\mu \longrightarrow A_\mu + \partial_\mu \epsilon. \tag{140}$$

Finally, the equations of motion of charged particles are given, in covariant form, by

$$m\frac{du^\mu}{d\tau} = eF^{\mu\nu}u_\nu, \tag{141}$$

where e is the charge of the particle and $u^\mu(\tau)$ its four-velocity as a function of the proper time.

The physical rôle of the vector potential becomes manifest only in Quantum Mechanics. Using the prescription of minimal substitution $\vec{p} \to \vec{p} - e\vec{A}$, the Schrödinger equation describing a particle with charge e moving in an electromagnetic field is

$$i\partial_t \Psi = \left[-\frac{1}{2m}\left(\vec{\nabla} - ie\vec{A}\right)^2 + e\varphi \right] \Psi. \tag{142}$$

Because of the explicit dependence on the electromagnetic potentials φ and $\vec{A}$, this equation seems to change under the gauge transformations (138). This is physically acceptable only if the ambiguity does not affect the probability density given by $|\Psi(t,\vec{x})|^2$. Therefore, a gauge transformation of the electromagnetic potential should amount to a change in the (unobservable) phase of the wave function. This is indeed what happens: the Schrödinger equation (142) is invariant under the gauge transformations (138) provided the phase of the wave function is transformed at the same time according to

$$\Psi(t,\vec{x}) \longrightarrow e^{-ie\,\epsilon(t,\vec{x})}\Psi(t,\vec{x}). \tag{143}$$

Aharonov-Bohm effect. This interplay between gauge transformations and the phase of the wave function give rise to surprising phenomena. The first evidence of the rôle played by the electromagnetic potentials at the quantum level was pointed out by Yakir Aharonov and David Bohm [20]. Let us consider a double slit experiment as shown in Fig. 7, where we have placed a shielded solenoid just behind the first screen. Although the magnetic field is confined to the interior of the solenoid, the vector potential is nonvanishing also outside. Of course the value of $\vec{A}$ outside the solenoid is a pure gauge, i.e. $\vec{\nabla}\times\vec{A} = \vec{0}$, however because the region outside the solenoid is not simply connected the vector potential cannot be gauged to zero everywhere. If we denote by $\Psi_1^{(0)}$ and $\Psi_2^{(0)}$ the wave functions for each of the two electron beams in the absence of the solenoid, the total wave function once the magnetic field is switched on can be written as

$$\begin{aligned} \Psi &= e^{ie\int_{\Gamma_1}\vec{A}\cdot d\vec{x}}\Psi_1^{(0)} + e^{ie\int_{\Gamma_2}\vec{A}\cdot d\vec{x}}\Psi_2^{(0)} \\ &= e^{ie\int_{\Gamma_1}\vec{A}\cdot d\vec{x}}\left[\Psi_1^{(0)} + e^{ie\oint_{\Gamma}\vec{A}\cdot d\vec{x}}\Psi_2^{(0)}\right], \end{aligned} \tag{144}$$

where Γ_1 and Γ_2 are two curves surrounding the solenoid from different sides, and Γ is any closed loop surrounding it. Therefore the relative phase between the two beams gets an extra term depending on the value of the vector potential outside the solenoid as

$$U = \exp\left[ie\oint_\Gamma \vec{A}\cdot d\vec{x}\right]. \tag{145}$$

Because of the change in the relative phase of the electron wave functions, the presence of the vector potential becomes observable even if the electrons do not feel the magnetic field. If we perform the double-slit experiment when the magnetic field inside the solenoid is switched off we will observe the

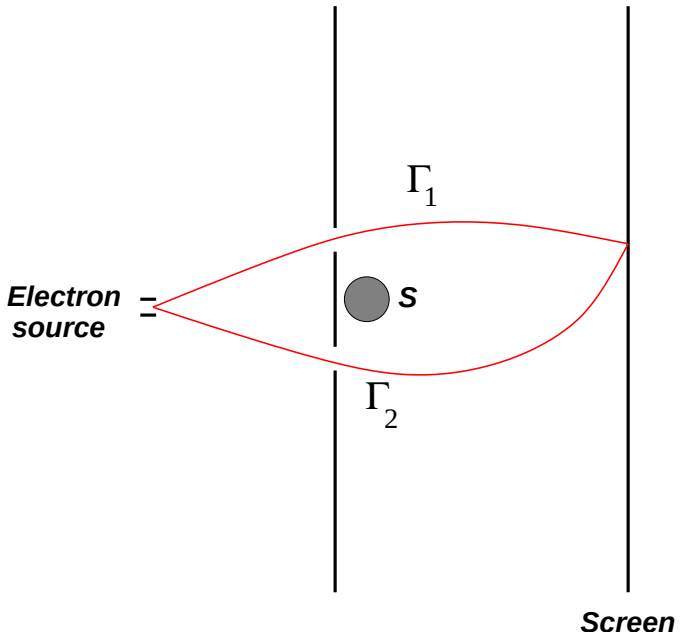


Fig. 7: Illustration of an interference experiment to show the Aharonov-Bohm effect. S represent the solenoid in whose interior the magnetic field is confined.

usual interference pattern on the second screen. However if now the magnetic field is switched on, because of the phase (144), a change in the interference pattern will appear. This is the Aharonov-Bohm effect.

The first question that comes up is what happens with gauge invariance. Since we said that $\vec{A}$ can be changed by a gauge transformation it seems that the resulting interference patters might depend on the gauge used. Actually, the phase U in (145) is independent of the gauge although, unlike other gauge-invariant quantities like $\vec{E}$ and $\vec{B}$, is nonlocal. Notice that, since $\vec{\nabla} \times \vec{A} = \vec{0}$ outside the solenoid, the value of U does not change under continuous deformations of the closed curve Γ, so long as it does not cross the solenoid.

The Dirac monopole. It is very easy to check that the vacuum Maxwell equations remain invariant under the transformation

$$\vec{E} - i\vec{B} \longrightarrow e^{i\theta}(\vec{E} - i\vec{B}), \qquad \theta \in [0, 2\pi] \tag{146}$$

which, in particular, for $\theta = \frac{\pi}{2}$ interchanges the electric and the magnetic fields: $\vec{E} \to \vec{B}$, $\vec{B} \to -\vec{E}$. This duality symmetry is however broken in the presence of electric sources. Nevertheless the Maxwell equations can be "completed" by introducing sources for the magnetic field $(\rho_m, \vec{\jmath}_m)$ in such a way that the duality (146) is restored when supplemented by the transformation

$$\rho - i\rho_m \longrightarrow e^{i\theta}(\rho - i\rho_m), \qquad \vec{\jmath} - i\vec{\jmath}_m \longrightarrow e^{i\theta}(\vec{\jmath} - i\vec{\jmath}_m). \tag{147}$$

Again for $\theta = \pi/2$ the electric and magnetic sources get interchanged.

In 1931 Dirac [21] studied the possibility of finding solutions of the completed Maxwell equation with a magnetic monopoles of charge g, i.e. solutions to

$$\vec{\nabla} \cdot \vec{B} = g\,\delta(\vec{x}). \tag{148}$$

Away from the position of the monopole $\vec{\nabla} \cdot \vec{B} = 0$ and the magnetic field can be still derived locally from a vector potential $\vec{A}$ according to $\vec{B} = \vec{\nabla} \times \vec{A}$. However, the vector potential cannot be regular

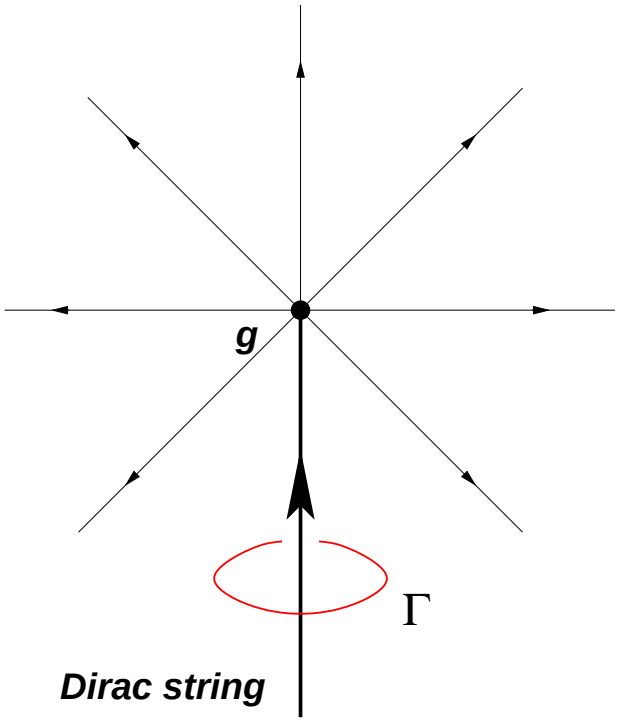


Fig. 8: The Dirac monopole.

everywhere since otherwise Gauss law would imply that the magnetic flux threading a closed surface around the monopole should vanish, contradicting (148).

We look now for solutions to Eq. (148). Working in spherical coordinates we find

$$B_r = \frac{g}{|\vec{x}|^2}, \qquad B_\varphi = B_\theta = 0. \tag{149}$$

Away from the position of the monopole ($\vec{x} \neq \vec{0}$) the magnetic field can be derived from the vector potential

$$A_\varphi = \frac{g}{|\vec{x}|} \tan\frac{\theta}{2}, \qquad A_r = A_\theta = 0. \tag{150}$$

As expected we find that this vector potential is actually singular around the half-line $\theta = \pi$ (see Fig. 8). This singular line starting at the position of the monopole is called the Dirac string and its position changes with a change of gauge but cannot be eliminated by any gauge transformation. Physically we can see it as an infinitely thin solenoid confining a magnetic flux entering into the magnetic monopole from infinity that equals the outgoing magnetic flux from the monopole.

Since the position of the Dirac string depends on the gauge chosen it seems that the presence of monopoles introduces an ambiguity. This would be rather strange, since Maxwell equations are gauge invariant also in the presence of magnetic sources. The solution to this apparent riddle lies in the fact that the Dirac string does not pose any consistency problem as far as it does not produce any physical effect, i.e. if its presence turns out to be undetectable. From our discussion of the Aharonov-Bohm effect we know that the wave function of charged particles pick up a phase (145) when surrounding a region where magnetic flux is confined (for example the solenoid in the Aharonov-Bohm experiment). As explained above, the Dirac string associated with the monopole can be seen as a infinitely thin solenoid. Therefore the Dirac string will be unobservable if the phase picked up by the wave function of a charged particle is equal to one. A simple calculation shows that this happens if

$$e^{i\,e\,g} = 1 \qquad \Longrightarrow \qquad e\,g = 2\pi n \;\text{ with }\; n \in \mathbb{Z}. \tag{151}$$

Interestingly, this discussion leads to the conclusion that the presence of a single magnetic monopoles somewhere in the Universe implies for consistency the quantization of the electric charge in units of $\frac{2\pi}{g}$, where g the magnetic charge of the monopole.

Quantization of the electromagnetic field. We now proceed to the quantization of the electromagnetic field in the absence of sources $\rho = 0$, $\vec{j} = \vec{0}$. In this case the Maxwell equations (139) can be derived from the Lagrangian density

$$\mathcal{L}_{\text{Maxwell}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} = \frac{1}{2}\left(\vec{E}^{\,2} - \vec{B}^{\,2}\right). \tag{152}$$

Although in general the procedure to quantize the Maxwell Lagrangian is not very different from the one used for the Klein-Gordon or the Dirac field, here we need to deal with a new ingredient: gauge invariance. Unlike the cases studied so far, here the photon field A_μ is not unambiguously defined because the action and the equations of motion are insensitive to the gauge transformations $A_\mu \to A_\mu + \partial_\mu \varepsilon$. A first consequence of this symmetry is that the theory has less physical degrees of freedom than one would expect from the fact that we are dealing with a vector field.

The way to tackle the problem of gauge invariance is to fix the freedom in choosing the electromagnetic potential before quantization. This can be done in several ways, for example by imposing the Lorentz gauge fixing condition

$$\partial_\mu A^\mu = 0. \tag{153}$$

Notice that this condition does not fix completely the gauge freedom since Eq. (153) is left invariant by gauge transformations satisfying $\partial_\mu \partial^\mu \varepsilon = 0$. One of the advantages, however, of the Lorentz gauge is that it is covariant and therefore does not pose any danger to the Lorentz invariance of the quantum theory. Besides, applying it to the Maxwell equation $\partial_\mu F^{\mu\nu} = 0$ one finds

$$0 = \partial_\mu \partial^\mu A^\nu - \partial_\nu \left(\partial_\mu A^\mu\right) = \partial_\mu \partial^\mu A^\nu, \tag{154}$$

which means that since A_μ satisfies the massless Klein-Gordon equation the photon, the quantum of the electromagnetic field, has zero mass.

Once gauge invariance is fixed A_μ is expanded in a complete basis of solutions to (154) and the canonical commutation relations are imposed

$$\widehat{A}_\mu(t,\vec{x}) = \sum_{\lambda=\pm 1} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2|\vec{k}|} \left[\epsilon_\mu(\vec{k},\lambda)\widehat{a}(\vec{k},\lambda) e^{-i|\vec{k}|t + i\vec{k}\cdot\vec{x}} + \epsilon_\mu(\vec{k},\lambda)^*\, \widehat{a}^\dagger(\vec{k},\lambda) e^{i|\vec{k}|t - i\vec{k}\cdot\vec{x}}\right] \tag{155}$$

where $\lambda = \pm 1$ represent the helicity of the photon, and $\epsilon_\mu(\vec{k},\lambda)$ are solutions to the equations of motion with well defined momentum an helicity. Because of (153) the polarization vectors have to be orthogonal to k_μ

$$k^\mu \epsilon_\mu(\vec{k},\lambda) = k^\mu \epsilon_\mu(\vec{k},\lambda)^* = 0. \tag{156}$$

The canonical commutation relations imply that

$$\begin{aligned} [\widehat{a}(\vec{k},\lambda), \widehat{a}^\dagger(\vec{k}\,',\lambda')] &= (2\pi)^3 (2|\vec{k}|)\delta(\vec{k} - \vec{k}\,')\delta_{\lambda\lambda'} \\ [\widehat{a}(\vec{k},\lambda), \widehat{a}(\vec{k}\,',\lambda')] &= [\widehat{a}^\dagger(\vec{k},\lambda), \widehat{a}^\dagger(\vec{k}\,',\lambda')] = 0. \end{aligned} \tag{157}$$

Therefore $\widehat{a}(\vec{k},\lambda)$, $\widehat{a}^\dagger(\vec{k},\lambda)$ form a set of creation-annihilation operators for photons with momentum $\vec{k}$ and helicity λ.

Behind the simple construction presented above there are a number of subleties related with gauge invariance. In particular the gauge freedom seem to introduce states in the Hilbert space with negative

probability. A careful analysis shows that when gauge invariance if properly handled these spurious states decouple from physical states and can be eliminated. The details can be found in standard textbooks [1]-[11].

Coupling gauge fields to matter. Once we know how to quantize the electromagnetic field we consider theories containing electrically charged particles, for example electrons. To couple the Dirac Lagrangian to electromagnetism we use as guiding principle what we learned about the Schrödinger equation for a charged particle. There we saw that the gauge ambiguity of the electromagnetic potential is compensated with a U(1) phase shift in the wave function. In the case of the Dirac equation we know that the Lagrangian is invariant under $\psi \to e^{ie\varepsilon}\psi$, with ε a constant. However this invariance is broken as soon as one identifies ε with the gauge transformation parameter of the electromagnetic field which depends on the position.

Looking at the Dirac Lagrangian (117) it is easy to see that in order to promote the global U(1) symmetry into a local one, $\psi \to e^{-ie\varepsilon(x)}\psi$, it suffices to replace the ordinary derivative ∂_μ by a covariant one D_μ satisfying

$$D_\mu \left[e^{-ie\varepsilon(x)}\psi\right] = e^{-ie\varepsilon(x)} D_\mu \psi. \tag{158}$$

This covariant derivative can be constructed in terms of the gauge potential A_μ as

$$D_\mu = \partial_\mu + ieA_\mu. \tag{159}$$

The Lagrangian of a spin-$\frac{1}{2}$ field coupled to electromagnetism is written as

$$\mathcal{L}_{\rm QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(i\not{D} - m)\psi, \tag{160}$$

invariant under the gauge transformations

$$\psi \longrightarrow e^{-ie\varepsilon(x)}\psi, \qquad A_\mu \longrightarrow A_\mu + \partial_\mu \varepsilon(x). \tag{161}$$

Unlike the theories we have seen so far, the Lagrangian (160) describe an interacting theory. By plugging (159) into the Lagrangian we find that the interaction between fermions and photons to be

$$\mathcal{L}^{(\rm int)}_{\rm QED} = -eA_\mu\, \overline{\psi}\gamma^\mu\psi. \tag{162}$$

As advertised above, in the Dirac theory the electric current four-vector is given by $j^\mu = e\overline{\psi}\gamma^\mu\psi$.

The quantization of interacting field theories poses new problems that we did not meet in the case of the free theories. In particular in most cases it is not possible to solve the theory exactly. When this happens the physical observables have to be computed in perturbation theory in powers of the coupling constant. An added problem appears when computing quantum corrections to the classical result, since in that case the computation of observables are plagued with infinities that should be taken care of. We will go back to this problem in section 8.

Nonabelian gauge theories. Quantum electrodynamics (QED) is the simplest example of a gauge theory coupled to matter based in the abelian gauge symmetry of local U(1) phase rotations. However, it is possible also to construct gauge theories based on nonabelian groups. Actually, our knowledge of the strong and weak interactions is based on the use of such nonabelian generalizations of QED.

Let us consider a gauge group G with generators T^a, $a = 1, \ldots, \dim G$ satisfying the Lie algebra[6]

$$[T^a, T^b] = if^{abc}T^c. \tag{163}$$

[6]Some basics facts about Lie groups have been summarized in Appendix A.

A gauge field taking values on the Lie algebra of $\mathcal{G}$ can be introduced $A_\mu \equiv A_\mu^a T^a$ which transforms under a gauge transformations as

$$A_\mu \longrightarrow -\frac{1}{ig} U \partial_\mu U^{-1} + U A_\mu U^{-1}, \qquad U = e^{i\chi^a(x) T^a}, \tag{164}$$

where g is the coupling constant. The associated field strength is defined as

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c. \tag{165}$$

Notice that this definition of the $F_{\mu\nu}^a$ reduces to the one used in QED in the abelian case when $f^{abc} = 0$. In general, however, unlike the case of QED the field strength is not gauge invariant. In terms of $F_{\mu\nu} = F_{\mu\nu}^a T^a$ it transforms as

$$F_{\mu\nu} \longrightarrow U F_{\mu\nu} U^{-1}. \tag{166}$$

The coupling of matter to a nonabelian gauge field is done by introducing again a covariant derivative. For a field in a representation of $\mathcal{G}$

$$\Phi \longrightarrow U\Phi \tag{167}$$

the covariant derivative is given by

$$D_\mu \Phi = \partial_\mu \Phi - i g A_\mu^a T^a \Phi. \tag{168}$$

With the help of this we can write a generic Lagrangian for a nonabelian gauge field coupled to scalars ϕ and spinors ψ as

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^a F^{\mu\nu\, a} + i\overline{\psi} \not{D} \psi + \overline{D_\mu \phi} D^\mu \phi - \overline{\psi} \left[M_1(\phi) + i\gamma_5 M_2(\phi) \right] \psi - V(\phi). \tag{169}$$

In order to keep the theory renormalizable we have to restrict $M_1(\phi)$ and $M_2(\phi)$ to be at most linear in ϕ whereas $V(\phi)$ have to be at most of quartic order. The Lagrangian of the standard model is of the form (169).

4.4 Understanding gauge symmetry

In classical mechanics the use of the Hamiltonian formalism starts with the replacement of generalized velocities by momenta

$$p_i \equiv \frac{\partial L}{\partial \dot{q}_i} \qquad \Longrightarrow \qquad \dot{q}_i = \dot{q}_i(q, p). \tag{170}$$

Most of the times there is no problem in inverting the relations $p_i = p_i(q, \dot{q})$. However in some systems these relations might not be invertible and result in a number of constraints of the type

$$f_a(q, p) = 0, \qquad a = 1, \ldots, N_1. \tag{171}$$

These systems are called degenerate or constrained [23, 24].

The presence of constraints of the type (171) makes the formulation of the Hamiltonian formalism more involved. The first problem is related to the ambiguity in defining the Hamiltonian, since the addition of any linear combination of the constraints do not modify its value. Secondly, one has to make sure that the constraints are consistent with the time evolution in the system. In the language of Poisson brackets this means that further constraints have to be imposed in the form

$$\{f_a, H\} \approx 0. \tag{172}$$

Following [23] we use the symbol $\approx$ to indicate a "weak" equality that holds when the constraints $f_a(q,p) = 0$ are satisfied. Notice however that since the computation of the Poisson brackets involves derivatives, the constraints can be used only after the bracket is computed. In principle the conditions (172) can give rise to a new set of constraints $g_b(q,p) = 0$, $b = 1, \ldots, N_2$. Again these constraints have to be consistent with time evolution and we have to repeat the procedure. Eventually this finishes when a set of constraints is found that do not require any further constraint to be preserved by the time evolution[7].

Once we find all the constraints of a degenerate system we consider the so-called first class constraints $\phi_a(q,p) = 0$, $a = 1, \ldots, M$, which are those whose Poisson bracket vanishes weakly

$$\{\phi_a, \phi_b\} = c_{abc}\phi_c \approx 0. \tag{173}$$

The constraints that do not satisfy this condition, called second class constraints, can be eliminated by modifying the Poisson bracket [23]. Then the total Hamiltonian of the theory is defined by

$$H_T = p_i q_i - L + \sum_{a=1}^{M} \lambda(t)\phi_a. \tag{174}$$

What has all this to do with gauge invariance? The interesting answer is that for a singular system the first class constraints ϕ_a generate gauge transformations. Indeed, because $\{\phi_a, \phi_b\} \approx 0 \approx \{\phi_a, H\}$ the transformations

$$\begin{aligned} q_i &\longrightarrow q_i + \sum_{a}^{M} \varepsilon_a(t)\{q_i, \phi_a\}, \\ p_i &\longrightarrow p_i + \sum_{a}^{M} \varepsilon_a(t)\{p_i, \phi_a\} \end{aligned} \tag{175}$$

leave invariant the state of the system. This ambiguity in the description of the system in terms of the generalized coordinates and momenta can be traced back to the equations of motion in Lagrangian language. Writing them in the form

$$\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}\ddot{q}_j = -\frac{\partial^2 L}{\partial \dot{q}_i \partial q_j}\dot{q}_j + \frac{\partial L}{\partial q_i}, \tag{176}$$

we find that order to determine the accelerations in terms of the positions and velocities the matrix $\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}$ has to be invertible. However, the existence of constraints (171) precisely implies that the determinant of this matrix vanishes and therefore the time evolution is not uniquely determined in terms of the initial conditions.

Let us apply this to Maxwell electrodynamics described by the Lagrangian

$$L = -\frac{1}{4}\int d^3\, F_{\mu\nu}F^{\mu\nu}. \tag{177}$$

The generalized momentum conjugate to A_μ is given by

$$\pi^\mu = \frac{\delta L}{\delta(\partial_0 A_\mu)} = F^{0\mu}. \tag{178}$$

In particular for the time component we find the constraint $\pi^0 = 0$. The Hamiltonian is given by

$$H = \int d^3x \left[\pi^\mu \partial_0 A_\mu - \mathcal{L}\right] = \int d^3x \left[\frac{1}{2}\left(\vec{E}^{\,2} + \vec{B}^{\,2}\right) + \pi^0\partial_0 A_0 + A_0\vec{\nabla}\cdot\vec{E}\right]. \tag{179}$$

[7] In principle it is also possible that the procedure finishes because some kind of inconsistent identity is found. In this case the system itself is inconsistent as it is the case with the Lagrangian $L(q,\dot{q}) = q$.

Requiring the consistency of the constraint $\pi^0 = 0$ we find a second constraint

$$\{\pi^0, H\} \approx \partial_0 \pi^0 + \vec{\nabla} \cdot \vec{E} = 0. \tag{180}$$

Together with the first constraint $\pi^0 = 0$ this one implies Gauss' law $\vec{\nabla} \cdot \vec{E} = 0$. These two constrains have vanishing Poisson bracket and therefore they are first class. Therefore the total Hamiltonian is given by

$$H_T = H + \int d^3x \left[\lambda_1(x)\pi^0 + \lambda_2(x)\vec{\nabla} \cdot \vec{E}\right], \tag{181}$$

where we have absorbed A_0 in the definition of the arbitrary functions $\lambda_1(x)$ and $\lambda_2(x)$. Actually, we can fix part of the ambiguity taking $\lambda_1 = 0$. Notice that, because A_0 has been included in the multipliers, fixing λ_1 amounts to fixing the value of A_0 and therefore it is equivalent to taking a temporal gauge. In this case the Hamiltonian is

$$H_T = \int d^3x \left[\frac{1}{2}\left(\vec{E}^{\,2} + \vec{B}^{\,2}\right) + \varepsilon(x)\vec{\nabla} \cdot \vec{E}\right] \tag{182}$$

and we are left just with Gauss' law as the only constraint. Using the canonical commutation relations

$$\{A_i(t, \vec{x}), E_j(t, \vec{x}\,')\} = \delta_{ij}\delta(\vec{x} - \vec{x}\,') \tag{183}$$

we find that the remaining gauge transformations are generated by Gauss' law

$$\delta A_i = \{A_i, \int d^3x' \, \varepsilon \, \vec{\nabla} \cdot \vec{E}\} = \partial_i \varepsilon, \tag{184}$$

while leaving A_0 invariant, so for consistency with the general gauge transformations the function $\varepsilon(x)$ should be independent of time. Notice that the constraint $\vec{\nabla} \cdot \vec{E} = 0$ can be implemented by demanding $\vec{\nabla} \cdot \vec{A} = 0$ which reduces the three degrees of freedom of $\vec{A}$ to the two physical degrees of freedom of the photon.

So much for the classical analysis. In the quantum theory the constraint $\vec{\nabla} \cdot \vec{E} = 0$ has to be imposed on the physical states $|\mathrm{phys}\rangle$. This is done by defining the following unitary operator on the Hilbert space

$$\mathcal{U}(\varepsilon) \equiv \exp\left(i \int d^3x \, \varepsilon(\vec{x}) \, \vec{\nabla} \cdot \vec{E}\right). \tag{185}$$

By definition, physical states should not change when a gauge transformations is performed. This is implemented by requiring that the operator $\mathcal{U}(\varepsilon)$ acts trivially on a physical state

$$\mathcal{U}(\varepsilon)|\mathrm{phys}\rangle = |\mathrm{phys}\rangle \qquad \Longrightarrow \qquad (\vec{\nabla} \cdot \vec{E})|\mathrm{phys}\rangle = 0. \tag{186}$$

In the presence of charge density ρ, the condition that physical states are annihilated by Gauss' law changes to $(\vec{\nabla} \cdot \vec{E} - \rho)|\mathrm{phys}\rangle = 0$.

The role of gauge transformations in the quantum theory is very illuminating in understanding the real rôle of gauge invariance [25]. As we have learned, the existence of a gauge symmetry in a theory reflects a degree of redundancy in the description of physical states in terms of the degrees of freedom appearing in the Lagrangian. In Classical Mechanics, for example, the state of a system is usually determined by the value of the canonical coordinates (q_i, p_i). We know, however, that this is not the case for constrained Hamiltonian systems where the transformations generated by the first class constraints change the value of q_i and p_i withoug changing the physical state. In the case of Maxwell theory for every physical configuration determined by the gauge invariant quantities $\vec{E}$, $\vec{B}$ there is an infinite number of possible values of the vector potential that are related by gauge transformations $\delta A_\mu = \partial_\mu \varepsilon$.

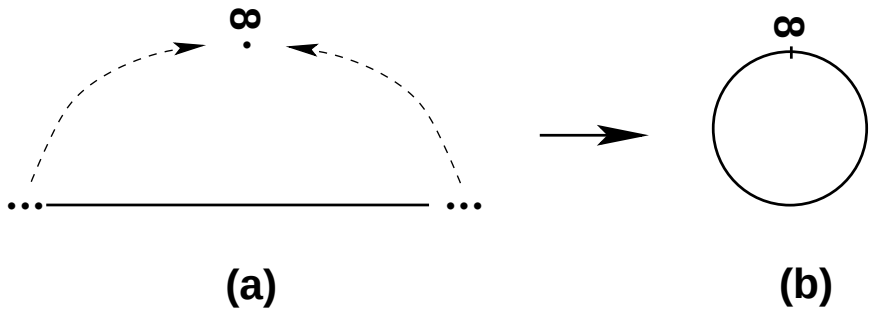


Fig. 9: Compactification of the real line (a) into the circumference S^1 (b) by adding the point at infinity.

In the quantum theory this means that the Hilbert space of physical states is defined as the result of identifying all states related by the operator $\mathcal{U}(\varepsilon)$ with any gauge function $\varepsilon(x)$ into a single physical state $|\mathrm{phys}\rangle$. In other words, each physical state corresponds to a whole orbit of states that are transformed among themselves by gauge transformations.

This explains the necessity of gauge fixing. In order to avoid the redundancy in the states a further condition can be given that selects one single state on each orbit. In the case of Maxwell electrodynamics the conditions $A_0 = 0$, $\vec{\nabla} \cdot \vec{A} = 0$ selects a value of the gauge potential among all possible ones giving the same value for the electric and magnetic fields.

Since states have to be identified by gauge transformations the topology of the gauge group plays an important physical rôle. To illustrate the point let us first deal with a toy model of a U(1) gauge theory in 1+1 dimensions. Later we will be more general. In the Hamiltonian formalism gauge transformations $g(\vec{x})$ are functions defined on $\mathbb{R}$ with values on the gauge group U(1)

$$g : \mathbb{R} \longrightarrow U(1). \tag{187}$$

We assume that $g(x)$ is regular at infinity. In this case we can add to the real line $\mathbb{R}$ the point at infinity to compactify it into the circumference S^1 (see Fig. 9). Once this is done $g(x)$ are functions defined on S^1 with values on $U(1) = S^1$ that can be parametrized as

$$g : S^1 \longrightarrow U(1), \qquad g(x) = e^{i\alpha(x)}, \tag{188}$$

with $x \in [0, 2\pi]$.

Because S^1 does have a nontrivial topology, $g(x)$ can be divided into topological sectors. These sectors are labelled by an integer number $n \in \mathbb{Z}$ and are defined by

$$\alpha(2\pi) = \alpha(0) + 2\pi\, n \ . \tag{189}$$

Geometrically n gives the number of times that the spatial S^1 winds around the S^1 defining the gauge group U(1). This winding number can be written in a more sophisticated way as

$$\oint_{S^1} g(x)^{-1} dg(x) = 2\pi n \ , \tag{190}$$

where the integral is along the spatial S^1.

In $\mathbb{R}^3$ a similar situation happens with the gauge group[8] SU(2). If we demand $g(\vec{x}) \in$ SU(2) to be regular at infinity $|\vec{x}| \to \infty$ we can compactify $\mathbb{R}^3$ into a three-dimensional sphere S^3, exactly as we did in 1+1 dimensions. On the other hand, the function $g(\vec{x})$ can be written as

$$g(\vec{x}) = a^0(x)\mathbf{1} + \vec{a}(x) \cdot \vec{\sigma} \tag{191}$$

[8] Although we present for simplicity only the case of SU(2), similar arguments apply to any simple group.

and the conditions $g(x)^\dagger g(x) = \mathbf{1}$, $\det g = 1$ implies that $(a^0)^2 + \vec{a}^{\,2} = 1$. Therefore SU(2) is a three-dimensional sphere and $g(x)$ defines a function

$$g : S^3 \longrightarrow S^3. \tag{192}$$

As it was the case in 1+1 dimensions here the gauge transformations $g(x)$ are also divided into topological sectors labelled this time by the winding number

$$n = \frac{1}{24\pi^2}\int_{S^3} d^3x\, \epsilon_{ijk}\mathrm{Tr}\,\left[\left(g^{-1}\partial_i g\right)\left(g^{-1}\partial_i g\right)\left(g^{-1}\partial_i g\right)\right] \in \mathbb{Z}. \tag{193}$$

In the two cases analyzed we find that due to the nontrivial topology of the gauge group manifold the gauge transformations are divided into different sectors labelled by an integer n. Gauge transformations with different values of n cannot be smoothly deformed into each other. The sector with $n = 0$ corresponds to those gauge transformations that can be connected with the identity.

Now we can be a bit more formal. Let us consider a gauge theory in 3+1 dimensions with gauge group G and let us denote by $\mathcal{G}$ the set of all gauge transformations $\mathcal{G} = \{g : S^3 \to G\}$. At the same time we define $\mathcal{G}_0$ as the set of transformations in $\mathcal{G}$ that can be smoothly deformed into the identity. Our theory will have topological sectors if

$$\mathcal{G}/\mathcal{G}_0 \neq \mathbf{1}. \tag{194}$$

In the case of the electromagnetism we have seen that Gauss' law annihilates physical states. For a nonabelian theory the analysis is similar and leads to the condition

$$\mathcal{U}(g_0)|\mathrm{phys}\rangle \equiv \exp\left[i\int d^3x\, \chi^a(\vec{x})\vec{\nabla}\cdot\vec{E}^a\right]|\mathrm{phys}\rangle = |\mathrm{phys}\rangle, \tag{195}$$

where $g_0(\vec{x}) = e^{i\chi^a(\vec{x})T^a}$ is in the connected component of the identity $\mathcal{G}_0$. The important point to realize here is that only the elements of $\mathcal{G}_0$ can be written as exponentials of the infinitesimal generators. Since this generators annihilate the physical states this implies that $\mathcal{U}(g_0)|\mathrm{phys}\rangle = |\mathrm{phys}\rangle$ only when $g_0 \in \mathcal{G}_0$.

What happens then with the other topological sectors? If $g \in \mathcal{G}/\mathcal{G}_0$ there is still a unitary operator $\mathcal{U}(g)$ that realizes gauge transformations on the Hilbert space of the theory. However since g is not in the connected component of the identity, it cannot be written as the exponential of Gauss' law. Still gauge invariance is preserved if $\mathcal{U}(g)$ only changes the overall global phase of the physical states. For example, if g_1 is a gauge transformation with winding number $n = 1$

$$\mathcal{U}(g_1)|\mathrm{phys}\rangle = e^{i\theta}|\mathrm{phys}\rangle. \tag{196}$$

It is easy to convince oneself that all transformations with winding number $n = 1$ have the same value of θ modulo 2π. This can be shown by noticing that if $g(\vec{x})$ has winding number $n = 1$ then $g(\vec{x})^{-1}$ has opposite winding number $n = -1$. Since the winding number is additive, given two transformations g_1, g_2 with winding number 1, $g_1^{-1}g_2$ has winding number $n = 0$. This implies that

$$|\mathrm{phys}\rangle = \mathcal{U}(g_1^{-1}g_2)|\mathrm{phys}\rangle = \mathcal{U}(g_1)^\dagger\mathcal{U}(g_2)|\mathrm{phys}\rangle = e^{i(\theta_2-\theta_1)}|\mathrm{phys}\rangle \tag{197}$$

and we conclude that $\theta_1 = \theta_2 \bmod 2\pi$. Once we know this it is straightforward to conclude that a gauge transformation $g_n(\vec{x})$ with winding number n has the following action on physical states

$$\mathcal{U}(g_n)|\mathrm{phys}\rangle = e^{in\theta}|\mathrm{phys}\rangle, \qquad n \in \mathbb{Z}. \tag{198}$$

To find a physical interpretation of this result we are going to look for similar things in other physical situations. One of then is borrowed from condensed matter physics and refers to the quantum

states of electrons in the periodic potential produced by the ion lattice in a solid. For simplicity we discuss the one-dimensional case where the minima of the potential are separated by a distance a. When the barrier between consecutive degenerate vacua is high enough we can neglect tunneling between different vacua and consider the ground state $|na\rangle$ of the potential near the minimum located at $x = na$ ($n \in \mathbb{Z}$) as possible vacua of the theory. This vacuum state is, however, not invariant under lattice translations

$$e^{ia\widehat{P}}|na\rangle = |(n+1)a\rangle. \tag{199}$$

However, it is possible to define a new vacuum state

$$|k\rangle = \sum_{n\in\mathbb{Z}} e^{-ikna}|na\rangle, \tag{200}$$

which under $e^{ia\widehat{P}}$ transforms by a global phase

$$e^{ia\widehat{P}}|k\rangle = \sum_{n\in\mathbb{Z}} e^{-ikna}|(n+1)a\rangle = e^{ika}|k\rangle. \tag{201}$$

This ground state is labelled by the momentum k and corresponds to the Bloch wave function.

This looks very much the same as what we found for nonabelian gauge theories. The vacuum state labelled by θ plays a rôle similar to the Bloch wave function for the periodic potential with the identification of θ with the momentum k. To make this analogy more precise let us write the Hamiltonian for nonabelian gauge theories

$$H = \frac{1}{2}\int d^3x \left(\vec{\pi}_a\cdot\vec{\pi}_a + \vec{B}_a\cdot\vec{B}_a\right) = \frac{1}{2}\int d^3x \left(\vec{E}_a\cdot\vec{E}_a + \vec{B}_a\cdot\vec{B}_a\right), \tag{202}$$

where we have used the expression of the canonical momenta π_a^i and we assume that the Gauss' law constraint is satisfied. Looking at this Hamiltonian we can interpret the first term within the brackets as the kinetic energy $T = \frac{1}{2}\vec{\pi}_a\cdot\vec{\pi}_a$ and the second term as the potential energy $V = \frac{1}{2}\vec{B}_a\cdot\vec{B}_a$. Since $V \geq 0$ we can identify the vacua of the theory as those $\vec{A}$ for which $V = 0$, modulo gauge transformations. This happens wherever $\vec{A}$ is a pure gauge. However, since we know that the gauge transformations are labelled by the winding number we can have an infinite number of vacua which cannot be continuously connected with one another using trivial gauge transformations. Taking a representative gauge transformation $g_n(\vec{x})$ in the sector with winding number n, these vacua will be associated with the gauge potentials

$$\vec{A} = -\frac{1}{ig}g_n(\vec{x})\vec{\nabla}g_n(\vec{x})^{-1}, \tag{203}$$

modulo topologically trivial gauge transformations. Therefore the theory is characterized by an infinite number of vacua $|n\rangle$ labelled by the winding number. These vacua are not gauge invariant. Indeed, a gauge transformation with $n = 1$ will change the winding number of the vacua in one unit

$$\mathcal{U}(g_1)|n\rangle = |n+1\rangle. \tag{204}$$

Nevertheless a gauge invariant vacuum can be defined as

$$|\theta\rangle = \sum_{n\in\mathbb{Z}} e^{-in\theta}|n\rangle, \qquad \text{with } \theta\in\mathbb{R} \tag{205}$$

satisfying

$$\mathcal{U}(g_1)|\theta\rangle = e^{i\theta}|\theta\rangle. \tag{206}$$

We have concluded that the nontrivial topology of the gauge group have very important physical consequences for the quantum theory. In particular it implies an ambiguity in the definition of the vacuum. Actually, this can also be seen in a Lagrangian analysis. In constructing the Lagrangian for the nonabelian version of Maxwell theory we only consider the term $F^a_{\mu\nu}F^{\mu\nu\,a}$. However this is not the only Lorentz and gauge invariant term that contains just two derivatives. We can write the more general Lagrangian

$$\mathcal{L} \;=\; -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu\,a} - \frac{\theta g^2}{32\pi^2}F^a_{\mu\nu}\widetilde{F}^{\mu\nu\,a}, \tag{207}$$

where $\widetilde{F}^a_{\mu\nu}$ is the dual of the field strength defined by

$$\widetilde{F}^a_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\sigma\lambda}F^{\sigma\lambda}. \tag{208}$$

The extra term in (207), proportional to $\vec{E}^{\,a}\cdot\vec{B}^{\,a}$, is actually a total derivative and does not change the equations of motion or the quantum perturbation theory. Nevertheless it has several important physical consequences. One of them is that it violates both parity P and the combination of charge conjugation and parity CP. This means that since strong interactions are described by a nonabelian gauge theory with group SU(3) there is an extra source of CP violation which puts a strong bound on the value of θ. One of the consequences of a term like (207) in the QCD Lagrangian is a nonvanishing electric dipole moment for the neutron [26]. The fact that this is not observed impose a very strong bound on the value of the θ-parameter

$$|\theta| < 10^{-9} \tag{209}$$

From a theoretical point of view it is still to be fully understood why θ either vanishes or has a very small value.

Finally, the θ-vacuum structure of gauge theories that we found in the Hamiltonian formalism can be also obtained using path integral techniques form the Lagrangian (207). The second term in Eq. (207) gives then a contribution that depends on the winding number of the corresponding gauge configuration.

5 Towards computational rules: Feynman diagrams

As the basic tool to describe the physics of elementary particles, the final aim of quantum field theory is the calculation of observables. Most of the information we have about the physics of subatomic particles comes from scattering experiments. Typically, these experiments consist of arranging two or more particles to collide with a certain energy and to setup an array of detectors, sufficiently far away from the region where the collision takes place, that register the outgoing products of the collision and their momenta (together with other relevant quantum numbers).

Next we discuss how these cross sections can be computed from quantum mechanical amplitudes and how these amplitudes themselves can be evaluated in perturbative quantum field theory. We keep our discussion rather heuristic and avoid technical details that can be found in standard texts [2]- [11]. The techniques described will be illustrated with the calculation of the cross section for Compton scattering at low energies.

5.1 Cross sections and S-matrix amplitudes

In order to fix ideas let us consider the simplest case of a collision experiment where two particles collide to produce again two particles in the final state. The aim of such an experiments is a direct measurement of the number of particles per unit time $\frac{dN}{dt}(\theta,\varphi)$ registered by the detector flying within a solid angle $d\Omega$ in the direction specified by the polar angles θ, φ (see Fig. 10). On general grounds we know that

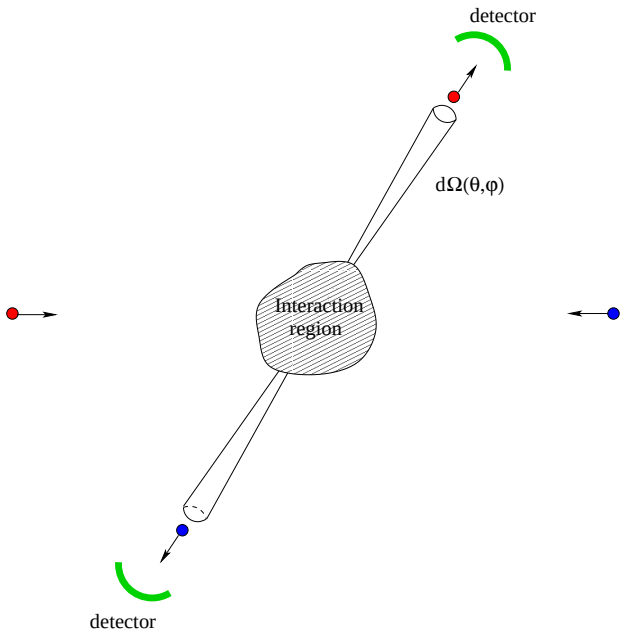


Fig. 10: Schematic setup of a two-to-two-particles single scattering event in the center of mass reference frame.

this quantity has to be proportional to the flux of incoming particles[9], $f_{\rm in}$. The proportionality constant defines the differential cross section

$$\frac{dN}{dt}(\theta,\varphi) = f_{\rm in}\frac{d\sigma}{d\Omega}(\theta,\varphi). \tag{210}$$

In natural units $f_{\rm in}$ has dimensions of (length)$^{-3}$, and then the differential cross section has dimensions of (length)^2. It depends, apart from the direction (θ,φ), on the parameters of the collision (energy, impact parameter, etc.) as well as on the masses and spins of the incoming particles.

Differential cross sections measure the angular distribution of the products of the collision. It is also physically interesting to quantify how effective the interaction between the particles is to produce a nontrivial dispersion. This is measured by the total cross section, which is obtained by integrating the differential cross section over all directions

$$\sigma = \int_{-1}^{1} d(\cos\theta)\int_0^{2\pi} d\varphi\,\frac{d\sigma}{d\Omega}(\theta,\varphi). \tag{211}$$

To get some physical intuition of the meaning of the total cross section we can think of the classical scattering of a point particle off a sphere of radius R. The particle undergoes a collision only when the impact parameter is smaller than the radius of the sphere and a calculation of the total cross section yields $\sigma = \pi R^2$. This is precisely the cross area that the sphere presents to incoming particles.

In Quantum Mechanics in general and in quantum field theory in particular the starting point for the calculation of cross sections is the probability amplitude for the corresponding process. In a scattering experiment one prepares a system with a given number of particles with definite momenta $\vec{p}_1,\ldots,\vec{p}_n$. In the Heisenberg picture this is described by a time independent state labelled by the incoming momenta of the particles (to keep things simple we consider spinless particles) that we denote by

$$|\vec{p}_1,\ldots,\vec{p}_n;{\rm in}\rangle. \tag{212}$$

[9]This is defined as the number of particles that enter the interaction region per unit time and per unit area perpendicular to the direction of the beam.

On the other hand, as a result of the scattering experiment a number k of particles with momenta $\vec{p}_1{}', \ldots, \vec{p}_k{}'$ are detected. Thus, the system is now in the "out" Heisenberg picture state

$$|\vec{p}_1{}', \ldots, \vec{p}_k{}'; \text{out}\rangle \tag{213}$$

labelled by the momenta of the particles detected at late times. The probability amplitude of detecting k particles in the final state with momenta $\vec{p}_1{}', \ldots, \vec{p}_k{}'$ in the collision of n particles with initial momenta $\vec{p}_1, \ldots, \vec{p}_n$ defines the S-matrix amplitude

$$S(\text{in} \to \text{out}) = \langle \vec{p}_1{}', \ldots, \vec{p}_k{}'; \text{out} | \vec{p}_1, \ldots, \vec{p}_n; \text{in}\rangle. \tag{214}$$

It is very important to keep in mind that both the (212) and (213) are time-independent states in the Hilbert space of a very complicated interacting theory. However, since both at early and late times the incoming and outgoing particles are well apart from each other, the "in" and "out" states can be thought as two states $|\vec{p}_1, \ldots, \vec{p}_n\rangle$ and $|\vec{p}_1{}', \ldots, \vec{p}_k{}'\rangle$ of the Fock space of the corresponding free theory in which the coupling constants are zero. Then, the overlaps (214) can be written in terms of the matrix elements of an S-matrix operator $\widehat{S}$ acting on the free Fock space

$$\langle \vec{p}_1{}', \ldots, \vec{p}_k{}'; \text{out} | \vec{p}_1, \ldots, \vec{p}_n; \text{in}\rangle = \langle \vec{p}_1{}', \ldots, \vec{p}_k{}' | \widehat{S} | \vec{p}_1, \ldots, \vec{p}_n\rangle. \tag{215}$$

The operator $\widehat{S}$ is unitary, $\widehat{S}^\dagger = \widehat{S}^{-1}$, and its matrix elements are analytic in the external momenta.

In any scattering experiment there is the possibility that the particles do not interact at all and the system is left in the same initial state. Then it is useful to write the S-matrix operator as

$$\widehat{S} = \mathbf{1} + i\widehat{T}, \tag{216}$$

where $\mathbf{1}$ represents the identity operator. In this way, all nontrivial interactions are encoded in the matrix elements of the T-operator $\langle \vec{p}_1{}', \ldots, \vec{p}_k{}' | i\widehat{T} | \vec{p}_1, \ldots, \vec{p}_n\rangle$. Since momentum has to be conserved, a global delta function can be factored out from these matrix elements to define the invariant scattering amplitude $i\mathcal{M}$

$$\langle \vec{p}_1{}', \ldots, \vec{p}_k{}' | i\widehat{T} | \vec{p}_1, \ldots, \vec{p}_n\rangle = (2\pi)^4 \delta^{(4)}\left(\sum_{\text{initial}} p_i - \sum_{\text{final}} p'_f \right) i\mathcal{M}(\vec{p}_1, \ldots, \vec{p}_n; \vec{p}_1{}', \ldots, \vec{p}_k{}') \tag{217}$$

Total and differential cross sections can be now computed from the invariant amplitudes. Here we consider the most common situation in which two particles with momenta $\vec{p}_1$ and $\vec{p}_2$ collide to produce a number of particles in the final state with momenta $\vec{p}_i{}'$. In this case the total cross section is given by

$$\sigma = \frac{1}{(2\omega_{p_1})(2\omega_{p_2})|\vec{v}_{12}|} \int \left[\prod_{\substack{\text{final} \\ \text{states}}} \frac{d^3 p'_i}{(2\pi)^3} \frac{1}{2\omega_{p'_i}} \right] \left| \mathcal{M}_{i \to f} \right|^2 (2\pi)^4 \delta^{(4)}\left(p_1 + p_2 - \sum_{\substack{\text{final} \\ \text{states}}} p'_i \right), \tag{218}$$

where $\vec{v}_{12}$ is the relative velocity of the two scattering particles. The corresponding differential cross section can be computed by dropping the integration over the directions of the final momenta. We will use this expression later in Section 5.3 to evaluate the cross section of Compton scattering.

We seen how particle cross sections are determined by the invariant amplitude for the corresponding proccess, i.e. S-matrix amplitudes. In general, in quantum field theory it is not possible to compute exactly these amplitudes. However, in many physical situations it can be argued that interactions are weak enough to allow for a perturbative evaluation. In what follows we will describe how S-matrix elements can be computed in perturbation theory using Feynman diagrams and rules. These are very convenient bookkeeping techniques allowing both to keep track of all contributions to a process at a given order in perturbation theory, and computing the different contributions.

5.2 Feynman rules

The basic quantities to be computed in quantum field theory are vacuum expectation values of products of the operators of the theory. Particularly useful are time-ordered Green functions,

$$\langle\Omega|T\Big[\mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\Big]|\Omega\rangle, \tag{219}$$

where $|\Omega\rangle$ is the the ground state of the theory and the time ordered product is defined

$$T\Big[\mathcal{O}_i(x)\mathcal{O}_j(y)\Big] = \theta(x^0-y^0)\mathcal{O}_i(x)\mathcal{O}_j(y) + \theta(y^0-x^0)\mathcal{O}_j(y)\mathcal{O}_i(x). \tag{220}$$

The generalization to products with more than two operators is straightforward: operators are always multiplied in time order, those evaluated at earlier times always to the right. The interest of these kind of correlation functions lies in the fact that they can be related to S-matrix amplitudes through the so-called reduction formula. To keep our discussion as simple as possible we will not derived it or even write it down in full detail. Its form for different theories can be found in any textbook. Here it suffices to say that the reduction formula simply states that any S-matrix amplitude can be written in terms of the Fourier transform of a time-ordered correlation function. Morally speaking

$$\begin{gathered}\langle\vec{p}_1{}',\dots,\vec{p}_m{}';\mathrm{out}|\vec{p}_1,\dots,\vec{p}_n;\mathrm{in}\rangle \\ \Downarrow \\ \int d^4x_1\dots\int d^4y_n\langle\Omega|T\Big[\phi(x_1)^\dagger\dots\phi(x_m)^\dagger\phi(y_1)\dots\phi(y_n)\Big]|\Omega\rangle\, e^{ip_1{}'\cdot x_1}\dots e^{-ip_n\cdot y_n},\end{gathered} \tag{221}$$

where $\phi(x)$ is the field whose elementary excitations are the particles involved in the scattering.

The reduction formula reduces the problem of computing S-matrix amplitudes to that of evaluating time-ordered correlation functions of field operators. These quantities are easy to compute exactly in the free theory. For an interacting theory the situation is more complicated, however. Using path integrals, the vacuum expectation value of the time-ordered product of a number of operators can be expressed as

$$\langle\Omega|T\Big[\mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\Big]|\Omega\rangle = \frac{\displaystyle\int\mathscr{D}\phi\mathscr{D}\phi^\dagger\,\mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\,e^{iS[\phi,\phi^\dagger]}}{\displaystyle\int\mathscr{D}\phi\mathscr{D}\phi^\dagger\,e^{iS[\phi,\phi^\dagger]}}. \tag{222}$$

For an theory with interactions, neither the path integral in the numerator or in the denominator is Gaussian and they cannot be calculated exactly. However, Eq. (222) is still very useful. The action $S[\phi,\phi^\dagger]$ can be split into the free (quadratic) piece and the interaction part

$$S[\phi,\phi^\dagger] = S_0[\phi,\phi^\dagger] + S_{\mathrm{int}}[\phi,\phi^\dagger]. \tag{223}$$

All dependence in the coupling constants of the theory comes from the second piece. Expanding now $\exp[iS_{\mathrm{int}}]$ in power series of the coupling constant we find that each term in the series expansion of both the numerator and the denominator has the structure

$$\int\mathscr{D}\phi\mathscr{D}\phi^\dagger\Big[\dots\Big]e^{iS_0[\phi,\phi^\dagger]}, \tag{224}$$

where "..." denotes certain monomial of fields. The important point is that now the integration measure only involves the free action, and the path integral in (224) is Gaussian and therefore can be computed exactly. The same conclusion can be reached using the operator formalism. In this case the correlation function (219) can be expressed in terms of correlation functions of operators in the interaction picture. The advantage of using this picture is that the fields satisfy the free equations of motion and therefore

can be expanded in creation-annihilation operators. The correlations functions are then easily computed using Wick's theorem.

Putting together all the previous ingredients we can calculate S-matrix amplitudes in a perturbative series in the coupling constants of the field theory. This can be done using Feynman diagrams and rules, a very economical way to compute each term in the perturbative expansion of the S-matrix amplitude for a given process. We will not detail the the construction of Feynman rules but just present them heuristically.

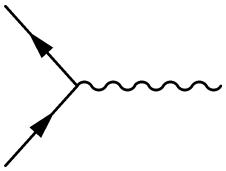
For the sake of concreteness we focus on the case of QED first. Going back to Eq. (160) we expand the covariant derivative to write the action

$$S_{\rm QED} = \int d^4x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \overline{\psi}(i\not{\partial} - m)\psi + e\overline{\psi}\gamma^\mu \psi A_\mu \right] . \tag{225}$$

The action contains two types of particles, photons and fermions, that we represent by straight and wavy lines respectively

The arrow in the fermion line does not represent the direction of the momentum but the flux of (negative) charge. This distinguishes particles form antiparticles: if the fermion propagates from left to right (i.e. in the direction of the charge flux) it represents a particle, whereas when it does from right to left it corresponds to an antiparticle. Photons are not charged and therefore wavy lines do not have orientation.

Next we turn to the interaction part of the action containing a photon field, a spinor and its conjugate. In a Feynman diagram this corresponds to the vertex



Now, in order to compute an S-matrix amplitude to a given order in the coupling constant e for a process with certain number of incoming and outgoing asymptotic states one only has to draw all possible diagrams with as many vertices as the order in perturbation theory, and the corresponding number and type of external legs. It is very important to keep in mind that in joining the fermion lines among the different building blocks of the diagram one has to respect their orientation. This reflects the conservation of the electric charge. In addition one should only consider diagrams that are topologically non-equivalent, i.e. that they cannot be smoothly deformed into one another keeping the external legs fixed[10].

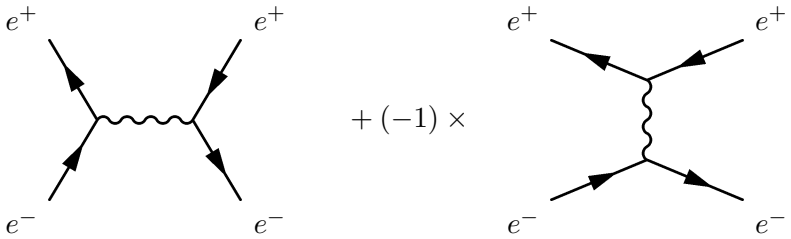
To show in a practical way how Feynman diagrams are drawn, we consider Bhabha scattering, i.e. the elastic dispersion of an electron and a positron:

$$e^+ + e^- \longrightarrow e^+ + e^- .$$

Our problem is to compute the S-matrix amplitude to the leading order in the electric charge. Because the QED vertex contains a photon line and our process does not have photons either in the initial or the

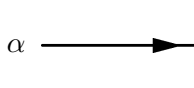
[10] From the point of view of the operator formalism, the requirement of considering only diagrams that are topologically nonequivalent comes from the fact that each diagram represents a certain Wick contraction in the correlation function of interaction-picture operators.

final states we find that drawing a Feynman diagram requires at least two vertices. In fact, the leading contribution is of order e^2 and comes from the following two diagrams, each containing two vertices:

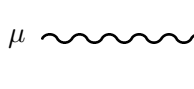


Incoming and outgoing particles appear respectively on the left and the right of this diagram. Notice how the identification of electrons and positrons is done comparing the direction of the charge flux with the direction of propagation. For electrons the flux of charges goes in the direction of propagation, whereas for positrons the two directions are opposite. These are the only two diagrams that can be drawn at this order in perturbation theory. It is important to include a relative minus sign between the two contributions. To understand the origin of this sign we have to remember that in the operator formalism Feynman diagrams are just a way to encode a particular Wick contraction of field operators in the interaction picture. The factor of -1 reflects the relative sign in Wick contractions represented by the two diagrams, due to the fermionic character of the Dirac field.

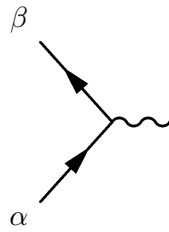
We have learned how to draw Feynman diagrams in QED. Now one needs to compute the contribution of each one to the corresponding amplitude using the so-called Feynman rules. The idea is simple: given a diagram, each of its building blocks (vertices as well as external and internal lines) has an associated contribution that allows the calculation of the corresponding diagram. In the case of QED in the Feynman gauge, we have the following correspondence for vertices and internal propagators:



$$\Longrightarrow \qquad \left(\frac{i}{\not{p} - m + i\varepsilon}\right)_{\beta\alpha}$$



$$\Longrightarrow \qquad \frac{-i\eta_{\mu\nu}}{p^2 + i\varepsilon}$$



$$\Longrightarrow \qquad -ie\gamma^{\mu}_{\beta\alpha}(2\pi)^4\delta^{(4)}(p_1 + p_2 + p_3).$$

A change in the gauge would reflect in an extra piece in the photon propagator. The delta function implementing conservation of momenta is written using the convention that all momenta are entering the vertex. In addition, one has to perform an integration over all momenta running in internal lines with the measure

$$\int \frac{d^d p}{(2\pi)^4}, \tag{226}$$

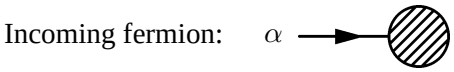
and introduce a factor of -1 for each fermion loop in the diagram[11].

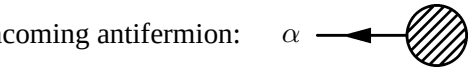
[11] The contribution of each diagram comes also multiplied by a degeneracy factor that takes into account in how many ways a given Wick contraction can be done. In QED, however, these factors are equal to 1 for many diagrams.

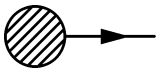
In fact, some of the integrations over internal momenta can actually be done using the delta function at the vertices, leaving just a global delta function implementing the total momentum conservation in the diagram [cf. Eq. (217)]. It is even possible that all integrations can be eliminated in this way. This is the case when we have tree level diagrams, i.e. those without closed loops. In the case of diagrams with loops there will be as many remaining integrations as the number of independent loops in the diagram.

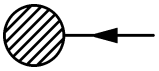
The need to perform integrations over internal momenta in loop diagrams has important consequences in Quantum Field Theory. The reason is that in many cases the resulting integrals are ill-defined, i.e. are divergent either at small or large values of the loop momenta. In the first case one speaks of *infrared divergences* and usually they cancel once all contributions to a given process are added together. More profound, however, are the divergences appearing at large internal momenta. These *ultraviolet divergences* cannot be cancelled and have to be dealt through the renormalization procedure. We will discuss this problem in some detail in Section 8.

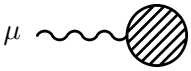
Were we computing time-ordered (amputated) correlation function of operators, this would be all. However, in the case of S-matrix amplitudes this is not the whole story. In addition to the previous rules here one needs to attach contributions also to the external legs in the diagram. These are the wave functions of the corresponding asymptotic states containing information about the spin and momenta of the incoming and outgoing particles. In the case of QED these contributions are:

Incoming fermion:  $\Longrightarrow \quad u_\alpha(\vec{p}, s)$

Incoming antifermion:  $\Longrightarrow \quad \overline{v}_\alpha(\vec{p}, s)$

Outgoing fermion:  $\Longrightarrow \quad \overline{u}_\alpha(\vec{p}, s)$

Outgoing antifermion:  $\Longrightarrow \quad v_\alpha(p, s)$

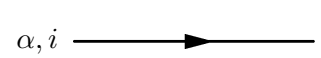
Incoming photon:  $\Longrightarrow \quad \epsilon_\mu(\vec{k}, \lambda)$

Outgoing photon:  $\Longrightarrow \quad \epsilon_\mu(\vec{k}, \lambda)^*$

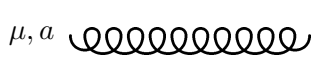
Here we have assumed that the momenta for incoming (resp. outgoing) particles are entering (resp. leaving) the diagram. It is important also to keep in mind that in the computation of S-matrix amplitudes all external states are on-shell. In Section 5.3 we illustrate the use of the Feynman rules for QED with the case of the Compton scattering.

The application of Feynman diagrams to carry out computations in perturbation theory is extremely convenient. It provides a very useful bookkeeping technique to account for all contributions to a process at a given order in the coupling constant. This does not mean that the calculation of Feynman diagrams is an easy task. The number of diagrams contributing to the process grows very fast with the order in perturbation theory and the integrals that appear in calculating loop diagrams also get very complicated. This means that, generically, the calculation of Feynman diagrams beyond the first few orders very often requires the use of computers.

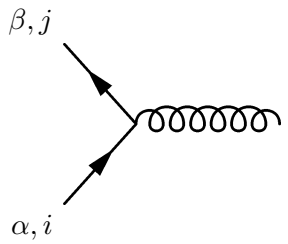
Above we have illustrated the Feynman rules with the case of QED. Similar rules can be computed for other interacting quantum field theories with scalar, vector or spinor fields. In the case of the nonabelian gauge theories introduced in Section 4.3 we have:



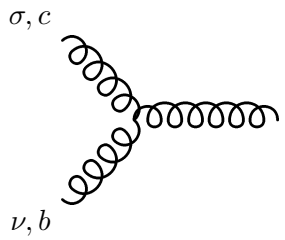
$$\Longrightarrow \qquad \left(\frac{i}{\not{p} - m + i\varepsilon}\right)_{\beta\alpha} \delta_{ij}$$



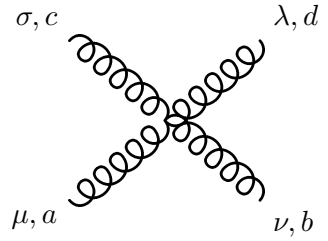
$$\Longrightarrow \qquad \frac{-i\eta_{\mu\nu}}{p^2 + i\varepsilon}\delta^{ab}$$



$$\Longrightarrow \qquad -ig\gamma^{\mu}_{\beta\alpha} t^a_{ij}$$



$$\Longrightarrow \qquad g\, f^{abc}\Big[\eta^{\mu\nu}(p_1^{\sigma} - p_2^{\sigma}) + \text{permutations}\Big]$$



$$\Longrightarrow \qquad -ig^2\Big[f^{abe}f^{cde}\Big(\eta^{\mu\sigma}\eta^{\nu\lambda} - \eta^{\mu\lambda}\eta^{\nu\sigma}\Big) + \text{permutations}\Big]$$

It is not our aim here to give a full and detailed description of the Feynman rules for nonabelian gauge theories. It suffices to point out that, unlike the case of QED, here the gauge fields can interact

among themselves. Indeed, the three and four gauge field vertices are a consequence of the cubic and quartic terms in the action

$$S = -\frac{1}{4}\int d^4x\, F^a_{\mu\nu}F^{\mu\nu\,a}, \tag{227}$$

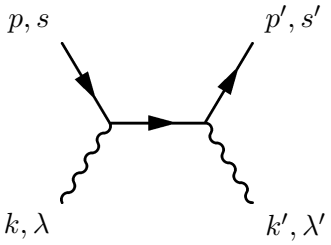
where the nonabelian gauge field strength $F^a_{\mu\nu}$ is given in Eq. (165). The self-interaction of the non-abelian gauge fields has crucial dynamical consequences and its at the very heart of its success in describing the physics of elementary particles.

5.3 An example: Compton scattering

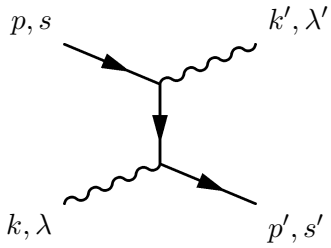
To illustrate the use of Feynman diagrams and Feynman rules we compute the cross section for the dispersion of photons by free electrons, the so-called Compton scattering:

$$\gamma(k,\lambda) + e^-(p,s) \longrightarrow \gamma(k',\lambda') + e^-(p',s').$$

In brackets we have indicated the momenta for the different particles, as well as the polarizations and spins of the incoming and outgoing photon and electrons respectively. The first step is to identify all the diagrams contributing to the process at leading order. Taking into account that the vertex of QED contains two fermion and one photon leg, it is straightforward to realize that any diagram contributing to the process at hand must contain at least two vertices. Hence the leading contribution is of order e^2. A first diagram we can draw is:

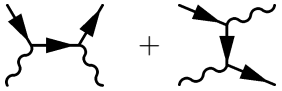


This is, however, not the only possibility. Indeed, there is a second possible diagram:



It is important to stress that these two diagrams are topologically nonequivalent, since deforming one into the other would require changing the label of the external legs. Therefore the leading $\mathcal{O}(e^2)$ amplitude has to be computed adding the contributions from both of them.

Using the Feynman rules of QED we find



$$\begin{aligned} &= (ie)^2\overline{u}(\vec{p}\,',s')\not{\epsilon}\,'(\vec{k}\,',\lambda')^*\frac{\not{p}+\not{k}+m_e}{(p+k)^2-m_e^2}\not{\epsilon}(\vec{k},\lambda)u(\vec{p},s) \\ &+ (ie)^2\overline{u}(\vec{p}\,',s')\not{\epsilon}(\vec{k},\lambda)\frac{\not{p}-\not{k}'+m_e}{(p-k')^2-m_e^2}\not{\epsilon}\,'(\vec{k}\,',\lambda')^*u(\vec{p},s). \end{aligned} \tag{228}$$

Because the leading order contributions only involve tree-level diagrams, there is no integration over internal momenta and therefore we are left with a purely algebraic expression for the amplitude. To get

an explicit expression we begin by simplifying the numerators. The following simple identity turns out to be very useful for this task

$$\not{a}\not{b} = -\not{b}\not{a} + 2(a \cdot b)\mathbf{1}. \tag{229}$$

Indeed, looking at the first term in Eq. (228) we have

$$\begin{aligned}(\not{p} + \not{k} + m_e)\not{\epsilon}(\vec{k},\lambda)u(\vec{p},s) &= -\not{\epsilon}(\vec{k},\lambda)(\not{p} - m_e)u(\vec{p},s) + \not{k}\not{\epsilon}(\vec{k},\lambda)u(\vec{p},s) \\ &+ 2p\cdot\epsilon(\vec{k},\lambda)u(\vec{p},s),\end{aligned} \tag{230}$$

where we have applied the identity (229) on the first term inside the parenthesis. The first term on the right-hand side of this equation vanishes identically because of Eq. (125). The expression can be further simplified if we restrict our attention to the Compton scattering at low energy when electrons are nonrelativistic. This means that all spatial momenta are much smaller than the electron mass

$$|\vec{p}|, |\vec{k}|, |\vec{p}\,'|, |\vec{k}\,'| \ll m_e. \tag{231}$$

In this approximation we have that $p^\mu, p'^\mu \approx (m_e, \vec{0})$ and therefore

$$p\cdot\epsilon(\vec{k},\lambda) = 0. \tag{232}$$

This follows from the absence of temporal photon polarization. Then we conclude that at low energies

$$(\not{p} + \not{k} + m_e)\not{\epsilon}(\vec{k},\lambda)u(\vec{p},s) = \not{k}\not{\epsilon}(\vec{k},\lambda)u(\vec{p},s) \tag{233}$$

and similarly for the second term in Eq. (228)

$$(\not{p} - \not{k}' + m_e)\not{\epsilon}'(\vec{k}',\lambda')^* u(\vec{p},s) = -\not{k}'\not{\epsilon}'(\vec{k}',\lambda')^* u(\vec{p},s). \tag{234}$$

Next, we turn to the denominators in Eq. (228). As it was explained in Section 5.2, in computing scattering amplitudes incoming and outgoing particles should have on-shell momenta,

$$p^2 = m_e^2 = p'^2 \qquad \text{and} \qquad k^2 = 0 = k'^2. \tag{235}$$

Then, the two denominator in Eq. (228) simplify respectively to

$$(p+k)^2 - m_e^2 = p^2 + k^2 + 2p\cdot k - m_e^2 = 2p\cdot k = 2\omega_p|\vec{k}| - 2\vec{p}\cdot\vec{k} \tag{236}$$

and

$$(p-k')^2 - m_e^2 = p^2 + k'^2 + 2p\cdot k' - m_e^2 = -2p\cdot k' = -2\omega_p|\vec{k}\,'| + 2\vec{p}\cdot\vec{k}\,'. \tag{237}$$

Working again in the low energy approximation (231) these two expressions simplify to

$$(p+k)^2 - m_e^2 \approx 2m_e|\vec{k}|, \qquad (p-k')^2 - m_e^2 \approx -2m_e|\vec{k}\,'|. \tag{238}$$

Putting together all these expressions we find that at low energies

$$\approx \frac{(ie)^2}{2m_e}\overline{u}(\vec{p}\,',s')\left[\not{\epsilon}'(\vec{k}\,'\lambda')^*\frac{\not{k}}{|\vec{k}|}\epsilon(\vec{k},\lambda) + \epsilon(\vec{k},\lambda)\frac{\not{k}'}{|\vec{k}\,'|}\not{\epsilon}'(\vec{k}\,'\lambda')^*\right]u(\vec{p},s). \tag{239}$$

Using now again the identity (229) a number of times as well as the transversality condition of the polarization vectors (156) we end up with a handier equation

$$\begin{aligned} &\approx \frac{e^2}{m_e}\Big[\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\Big]\overline{u}(\vec{p}\,',s')\frac{\not{k}}{|\vec{k}|}u(\vec{p},s) \\ &+ \frac{e^2}{2m_e}\overline{u}(\vec{p}\,',s')\not{\epsilon}(\vec{k},\lambda)\not{\epsilon}'(\vec{k}\,',\lambda')^*\left(\frac{\not{k}}{|\vec{k}|}-\frac{\not{k}'}{|\vec{k}\,'|}\right)u(\vec{p},s). \end{aligned} \tag{240}$$

With a little bit of effort we can show that the second term on the right-hand side vanishes. First we notice that in the low energy limit $|\vec{k}|\approx|\vec{k}\,'|$. If in addition we make use the conservation of momentum $k-k'=p'-p$ and the identity (125)

$$\begin{aligned} &\overline{u}(\vec{p}\,',s')\not{\epsilon}(\vec{k},\lambda)\not{\epsilon}'(\vec{k}\,',\lambda')^*\left(\frac{\not{k}}{|\vec{k}|}-\frac{\not{k}'}{|\vec{k}\,'|}\right)u(\vec{p},s) \\ &\qquad\approx \frac{1}{|\vec{k}|}\overline{u}(\vec{p}\,',s')\not{\epsilon}(\vec{k},\lambda)\not{\epsilon}'(\vec{k}\,',\lambda')^*(\not{p}'-m_e)u(\vec{p},s). \end{aligned} \tag{241}$$

Next we use the identity (229) to take the term $(\not{p}'-m_e)$ to the right. Taking into account that in the low energy limit the electron four-momenta are orthogonal to the photon polarization vectors [see Eq. (232)] we conclude that

$$\begin{aligned} &\overline{u}(\vec{p}\,',s')\not{\epsilon}(\vec{k},\lambda)\not{\epsilon}'(\vec{k}\,',\lambda')^*(\not{p}'-m_e)u(\vec{p},s) \\ &\qquad=\overline{u}(\vec{p}\,',s')(\not{p}'-m_e)\not{\epsilon}(\vec{k},\lambda)\not{\epsilon}'(\vec{k}\,',\lambda')^*u(\vec{p},s)=0 \end{aligned} \tag{242}$$

where the last identity follows from the equation satisfied by the conjugate positive-energy spinor, $\overline{u}(\vec{p}\,',s')(\not{p}'-m_e)=0$.

After all these lengthy manipulations we have finally arrived at the expression of the invariant amplitude for the Compton scattering at low energies

$$i\mathcal{M}=\frac{e^2}{m_e}\Big[\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\Big]\overline{u}(\vec{p}\,',s')\frac{\not{k}}{|\vec{k}|}u(\vec{p},s). \tag{243}$$

The calculation of the cross section involves computing the modulus squared of this quantity. For many physical applications, however, one is interested in the dispersion of photons with a given polarization by electrons that are not polarized, i.e. whose spins are randomly distributed. In addition in many situations either we are not interested, or there is no way to measure the final polarization of the outgoing electron. This is for example the situation in cosmology, where we do not have any information about the polarization of the free electrons in the primordial plasma before or after the scattering with photons (although we have ways to measure the polarization of the scattered photons).

To describe this physical situations we have to average over initial electron polarization (since we do not know them) and sum over all possible final electron polarization (because our detector is blind to this quantum number),

$$\overline{|i\mathcal{M}|^2}=\frac{1}{2}\left(\frac{e^2}{m_e|\vec{k}|}\right)^2\Big|\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\Big|^2\sum_{s=\pm\frac{1}{2}}\sum_{s'=\pm\frac{1}{2}}\Big|\overline{u}(\vec{p}\,',s')\not{k}u(\vec{p},s)\Big|^2. \tag{244}$$

The factor of $\frac{1}{2}$ comes from averaging over the two possible polarizations of the incoming electrons. The sums in this expression can be calculated without much difficulty. Expanding the absolute value explicitly

$$\sum_{s=\pm\frac{1}{2}}\sum_{s'=\pm\frac{1}{2}}\Big|\overline{u}(\vec{p}\,',s')\not{k}u(\vec{p},s)\Big|^2=\sum_{s=\pm\frac{1}{2}}\sum_{s'=\pm\frac{1}{2}}\Big[u(\vec{p},s)^\dagger\not{k}^\dagger\overline{u}(\vec{p}\,',s')^\dagger\Big]\Big[\overline{u}(\vec{p}\,',s')\not{k}u(\vec{p},s)\Big], \tag{245}$$

using that $\gamma^{\mu\dagger} = \gamma^0\gamma^\mu\gamma^0$ and after some manipulation one finds that

$$\sum_{s=\pm\frac{1}{2}}\sum_{s'=\pm\frac{1}{2}}\left|\overline{u}(\vec{p}\,',s')\not{k}u(\vec{p},s)\right|^2 = \left[\sum_{s=\pm\frac{1}{2}}u_\alpha(\vec{p},s)\overline{u}_\beta(\vec{p},s)\right](\not{k})_{\beta\sigma}\left[\sum_{s'=\pm\frac{1}{2}}u_\sigma(\vec{p}\,',s')\overline{u}_\rho(\vec{p}\,',s')\right](\not{k})_{\rho\alpha}$$
$$= \mathrm{Tr}\left[(\not{p}+m_e)\not{k}(\not{p}\,'+m_e)\not{k}\right], \tag{246}$$

where the final expression has been computed using the completeness relations in Eq. (128). The final evaluation of the trace can be done using the standard Dirac matrices identities. Here we compute it applying again the relation (229) to commute $\not{p}'$ and $\not{k}$. Using that $k^2 = 0$ and that we are working in the low energy limit we have[12]

$$\mathrm{Tr}\left[(\not{p}+m_e)\not{k}(\not{p}\,'+m_e)\not{k}\right] = 2(p\cdot k)(p'\cdot k)\mathrm{Tr}\,\mathbf{1} \approx 8m_e^2|\vec{k}|^2. \tag{247}$$

This gives the following value for the invariant amplitude

$$\overline{|i\mathcal{M}|^2} = 4e^4\left|\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\right|^2 \tag{248}$$

Plugging $\overline{|i\mathcal{M}|^2}$ into the formula for the differential cross section we get

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2m_e^2}\overline{|i\mathcal{M}|^2} = \left(\frac{e^2}{4\pi m_e}\right)^2\left|\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\right|^2. \tag{249}$$

The prefactor of the last equation is precisely the square of the so-called classical electron radius r_{cl}. In fact, the previous differential cross section can be rewritten as

$$\frac{d\sigma}{d\Omega} = \frac{3}{8\pi}\sigma_T\left|\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\right|^2, \tag{250}$$

where σ_T is the total Thomson cross section

$$\sigma_T = \frac{e^4}{6\pi m_e^2} = \frac{8\pi}{3}r_{\mathrm{cl}}^2. \tag{251}$$

The result (250) is relevant in many areas of Physics, but its importance is paramount in the study of the cosmological microwave background (CMB). Just before recombination the universe is filled by a plasma of electrons interacting with photons via Compton scattering, with temperatures of the order of 1 keV. Electrons are then nonrelativistic ($m_e \sim 0.5$ MeV) and the approximations leading to Eq. (250) are fully valid. Because we do not know the polarization state of the photons before being scattered by electrons we have to consider the cross section averaged over incoming photon polarizations. From Eq. (250) we see that this is proportional to

$$\frac{1}{2}\sum_{\lambda=1,2}\left|\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\right|^2 = \left[\frac{1}{2}\sum_{\lambda=1,2}\epsilon_i(\vec{k},\lambda)\epsilon_j(\vec{k},\lambda)^*\right]\epsilon_j(\vec{k}\,',\lambda')\epsilon_i(\vec{k}\,',\lambda')^*. \tag{252}$$

The sum inside the brackets can be computed using the normalization of the polarization vectors, $|\vec{\epsilon}(\vec{k},\lambda)|^2 = 1$, and the transversality condition $\vec{k}\cdot\vec{\epsilon}(\vec{k},\lambda) = 0$

$$\frac{1}{2}\sum_{\lambda=1,2}\left|\epsilon(\vec{k},\lambda)\cdot\epsilon'(\vec{k}\,',\lambda')^*\right|^2 = \frac{1}{2}\left(\delta_{ij}-\frac{k_ik_j}{|\vec{k}|^2}\right)\epsilon'_j(\vec{k}\,',\lambda')\epsilon'_i(\vec{k}\,',\lambda')^*$$

[12] We use also the fact that the trace of the product of an odd number of Dirac matrices is always zero.

$$= \frac{1}{2}\Big[1 - |\vec{\ell}\cdot\vec{\epsilon}\,'(\vec{k}\,',\lambda')|^2\Big], \tag{253}$$

where $\vec{\ell} = \frac{\vec{k}}{|\vec{k}|}$ is the unit vector in the direction of the incoming photon.

From the last equation we conclude that Thomson scattering suppresses all polarizations parallel to the direction of the incoming photon $\vec{\ell}$, whereas the differential cross section reaches the maximum in the plane normal to $\vec{\ell}$. If photons would collide with the electrons in the plasma with the same intensity from all directions, the result would be an unpolarized CMB radiation. The fact that polarization is actually measured in the CMB carries crucial information about the physics of the plasma before recombination and, as a consequence, about the very early universe (see for example [22] for a throughout discussion).

6 Symmetries

6.1 Noether's theorem

In Classical Mechanics and Classical Field Theory there is a basic result that relates symmetries and conserved charges. This is called Noether's theorem and states that for each continuous symmetry of the system there is conserved current. In its simplest version in Classical Mechanics it can be easily proved. Let us consider a Lagrangian $L(q_i, \dot{q}_i)$ which is invariant under a transformation $q_i(t) \to q_i'(t,\epsilon)$ labelled by a parameter ϵ. This means that $L(q', \dot{q}') = L(q, \dot{q})$ without using the equations of motion[13]. If $\epsilon \ll 1$ we can consider an infinitesimal variation of the coordinates $\delta_\epsilon q_i(t)$ and the invariance of the Lagrangian implies

$$0 = \delta_\epsilon L(q_i, \dot{q}_i) = \frac{\partial L}{\partial q_i}\delta_\epsilon q_i + \frac{\partial L}{\partial \dot{q}_i}\delta_\epsilon \dot{q}_i = \left[\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i}\right]\delta_\epsilon q_i + \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\delta_\epsilon q_i\right). \tag{254}$$

When $\delta_\epsilon q_i$ is applied on a solution to the equations of motion the term inside the square brackets vanishes and we conclude that there is a conserved quantity

$$\dot{Q} = 0 \quad \text{with} \quad Q \equiv \frac{\partial L}{\partial \dot{q}_i}\delta_\epsilon q_i. \tag{255}$$

Notice that in this derivation it is crucial that the symmetry depends on a continuous parameter since otherwise the infinitesimal variation of the Lagrangian in Eq. (254) does not make sense.

In Classical Field Theory a similar result holds. Let us consider for simplicity a theory of a single field $\phi(x)$. We say that the variations $\delta_\epsilon\phi$ depending on a continuous parameter ϵ are a symmetry of the theory if, without using the equations of motion, the Lagrangian density changes by

$$\delta_\epsilon \mathcal{L} = \partial_\mu K^\mu. \tag{256}$$

If this happens then the action remains invariant and so do the equations of motion. Working out now the variation of $\mathcal{L}$ under $\delta_\epsilon\phi$ we find

$$\partial_\mu K^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu\delta_\epsilon\phi + \frac{\partial\mathcal{L}}{\partial\phi}\delta_\epsilon\phi = \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta_\epsilon\phi\right) + \left[\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\right]\delta_\epsilon\phi. \tag{257}$$

If $\phi(x)$ is a solution to the equations of motion the last terms disappears, and we find that there is a conserved current

$$\partial_\mu J^\mu = 0 \quad \text{with} \quad J^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta_\epsilon\phi - K^\mu. \tag{258}$$

[13] The following result can be also derived a more general situations where the Lagrangian changes by a total time derivative.

Actually a conserved current implies the existence of a charge

$$Q \equiv \int d^3x\, J^0(t,\vec{x}) \tag{259}$$

which is conserved

$$\frac{dQ}{dt} = \int d^3x\, \partial_0 J^0(t,\vec{x}) = -\int d^3x\, \partial_i J^i(t,\vec{x}) = 0, \tag{260}$$

provided the fields vanish at infinity fast enough. Moreover, the conserved charge Q is a Lorentz scalar. After canonical quantization the charge Q defined by Eq. (259) is promoted to an operator that generates the symmetry on the fields

$$\delta\phi = i[\phi, Q]. \tag{261}$$

As an example we can consider a scalar field $\phi(x)$ which under a coordinate transformation $x \to x'$ changes as $\phi'(x') = \phi(x)$. In particular performing a space-time translation $x^{\mu\,\prime} = x^\mu + a^\mu$ we have

$$\phi'(x) - \phi(x) = -a^\mu \partial_\mu \phi + \mathcal{O}(a^2) \qquad \Longrightarrow \qquad \delta\phi = -a^\mu \partial_\mu \phi. \tag{262}$$

Since the Lagrangian density is also a scalar quantity, it transforms under translations as

$$\delta\mathcal{L} = -a^\mu \partial_\mu \mathcal{L}. \tag{263}$$

Therefore the corresponding conserved charge is

$$J^\mu = -\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} a^\nu \partial_\nu \phi + a^\mu \mathcal{L} \equiv -a_\nu T^{\mu\nu}, \tag{264}$$

where we introduced the energy-momentum tensor

$$T^{\mu\nu} = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial^\nu \phi - \eta^{\mu\nu}\mathcal{L}. \tag{265}$$

We find that associated with the invariance of the theory with respect to space-time translations there are four conserved currents defined by $T^{\mu\nu}$ with $\nu = 0, \ldots, 3$, each one associated with the translation along a space-time direction. These four currents form a rank-two tensor under Lorentz transformations satisfying

$$\partial_\mu T^{\mu\nu} = 0. \tag{266}$$

The associated conserved charges are given by

$$P^\nu = \int d^3x\, T^{0\nu} \tag{267}$$

and correspond to the total energy-momentum content of the field configuration. Therefore the energy density of the field is given by T^{00} while T^{0i} is the momentum density. In the quantum theory the P^μ are the generators of space-time translations.

Another example of a symmetry related with a physically relevant conserved charge is the global phase invariance of the Dirac Lagrangian (117), $\psi \to e^{i\theta}\psi$. For small θ this corresponds to variations $\delta_\theta\psi = i\theta\psi$, $\delta_\theta\overline{\psi} = -i\theta\overline{\psi}$ which by Noether's theorem result in the conserved charge

$$j^\mu = \overline{\psi}\gamma^\mu\psi, \qquad \partial_\mu j^\mu = 0. \tag{268}$$

Thus implying the existence of a conserved charge

$$Q = \int d^3x \overline{\psi}\gamma^0\psi = \int d^3x \psi^\dagger\psi. \tag{269}$$

In physics there are several instances of global U(1) symmetries that act as phase shifts on spinors. This is the case, for example, of the baryon and lepton number conservation in the standard model. A more familiar case is the U(1) local symmetry associated with electromagnetism. Notice that although in this case we are dealing with a local symmetry, $\theta \to e\alpha(x)$, the invariance of the Lagrangian holds in particular for global transformations and therefore there is a conserved current $j^\mu = e\overline{\psi}\gamma^\mu\psi$. In Eq. (162) we saw that the spinor is coupled to the photon field precisely through this current. Its time component is the electric charge density ρ, while the spatial components are the current density vector $\vec{j}$.

This analysis can be carried over also to nonabelian unitary global symmetries acting as

$$\psi_i \longrightarrow U_{ij}\psi_j, \qquad U^\dagger U = \mathbf{1} \tag{270}$$

and leaving invariant the Dirac Lagrangian when we have several fermions. If we write the matrix U in terms of the hermitian group generators T^a as

$$U = \exp\left(i\alpha_a T^a\right), \qquad (T^a)^\dagger = T^a, \tag{271}$$

we find the conserved current

$$j^{\mu\,a} = \overline{\psi}_i T^a_{ij}\gamma^\mu\psi_j, \qquad \partial_\mu j^\mu = 0. \tag{272}$$

This is the case, for example of the approximate flavor symmetries in hadron physics. The simplest example is the isospin symmetry that mixes the quarks u and d

$$\begin{pmatrix} u \\ d \end{pmatrix} \longrightarrow M \begin{pmatrix} u \\ d \end{pmatrix}, \qquad M \in \mathrm{SU}(2). \tag{273}$$

Since the proton is a bound state of two quarks u and one quark d while the neutron is made out of one quark u and two quarks d, this isospin symmetry reduces at low energies to the well known isospin transformations of nuclear physics that mixes protons and neutrons.

6.2 Symmetries in the quantum theory

We have seen that in canonical quantization the conserved charges Q^a associated to symmetries by Noether's theorem are operators implementing the symmetry at the quantum level. Since the charges are conserved they must commute with the Hamiltonian

$$[Q^a, H] = 0. \tag{274}$$

There are several possibilities in the quantum mechanical realization of a symmetry:

Wigner-Weyl realization. In this case the ground state of the theory $|0\rangle$ is invariant under the symmetry. Since the symmetry is generated by Q^a this means that

$$\mathcal{U}(\alpha)|0\rangle \equiv e^{i\alpha_a Q^a}|0\rangle = |0\rangle \quad \Longrightarrow \quad Q^a|0\rangle = 0. \tag{275}$$

At the same time the fields of the theory have to transform according to some irreducible representation of the group generated by the Q^a. From Eq. (261) it is easy to prove that

$$\mathcal{U}(\alpha)\phi_i\mathcal{U}(\alpha)^{-1} = U_{ij}(\alpha)\phi_j, \tag{276}$$

where $U_{ij}(\alpha)$ is an element of the representation in which the field ϕ_i transforms. If we consider now the quantum state associated with the operator ϕ_i

$$|i\rangle = \phi_i|0\rangle \tag{277}$$

we find that because of the invariance of the vacuum (275) the states $|i\rangle$ transform in the same representation as ϕ_i

$$\mathcal{U}(\alpha)|i\rangle = \mathcal{U}(\alpha)\phi_i\mathcal{U}(\alpha)^{-1}\mathcal{U}(\alpha)|0\rangle = U_{ij}(\alpha)\phi_j|0\rangle = U_{ij}(\alpha)|j\rangle. \tag{278}$$

Therefore the spectrum of the theory is classified in multiplets of the symmetry group. In addition, since $[H,\mathcal{U}(\alpha)] = 0$ all states in the same multiplet have the same energy. If we consider one-particle states, then going to the rest frame we conclude that all states in the same multiplet have exactly the same mass.

Nambu-Goldstone realization. In our previous discussion the result that the spectrum of the theory is classified according to multiplets of the symmetry group depended crucially on the invariance of the ground state. However this condition is not mandatory and one can relax it to consider theories where the vacuum state is not left invariant by the symmetry

$$e^{i\alpha_a Q^a}|0\rangle \neq |0\rangle \qquad \Longrightarrow \qquad Q^a|0\rangle \neq 0. \tag{279}$$

In this case it is also said that the symmetry is spontaneously broken by the vacuum.

To illustrate the consequences of (279) we consider the example of a number scalar fields φ^i ($i = 1, \ldots, N$) whose dynamics is governed by the Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi^i\partial^\mu\varphi^i - V(\varphi), \tag{280}$$

where we assume that $V(\phi)$ is bounded from below. This theory is globally invariant under the transformations

$$\delta\varphi^i = \epsilon^a(T^a)^i_j\varphi^j, \tag{281}$$

with T^a, $a = 1, \ldots, \frac{1}{2}N(N-1)$ the generators of the group SO(N).

To analyze the structure of vacua of the theory we construct the Hamiltonian

$$H = \int d^3x \left[\frac{1}{2}\pi^i\pi^i + \frac{1}{2}\vec{\nabla}\varphi^i\cdot\vec{\nabla}\varphi^i + V(\varphi)\right] \tag{282}$$

and look for the minimum of

$$\mathcal{V}(\varphi) = \int d^3x \left[\frac{1}{2}\vec{\nabla}\varphi^i\cdot\vec{\nabla}\varphi^i + V(\varphi)\right]. \tag{283}$$

Since we are interested in finding constant field configurations, $\vec{\nabla}\varphi = \vec{0}$ to preserve translational invariance, the vacua of the potential $\mathcal{V}(\varphi)$ coincides with the vacua of $V(\varphi)$. Therefore the minima of the potential correspond to the vacuum expectation values[14]

$$\langle\varphi^i\rangle: \qquad V(\langle\varphi^i\rangle) = 0, \qquad \left.\frac{\partial V}{\partial\varphi^i}\right|_{\varphi^i=\langle\varphi^i\rangle} = 0. \tag{284}$$

We divide the generators T^a of SO(N) into two groups: Those denoted by H^α ($\alpha = 1, \ldots, h$) that satisfy

$$(H^\alpha)^i_j\langle\varphi^j\rangle = 0. \tag{285}$$

[14]For simplicity we consider that the minima of $V(\phi)$ occur at zero potential.

This means that the vacuum configuration $\langle\varphi^i\rangle$ is left invariant by the transformation generated by H^α. For this reason we call them *unbroken generators*. Notice that the commutator of two unbroken generators also annihilates the vacuum expectation value, $[H^\alpha, H^\beta]_{ij}\langle\varphi^j\rangle = 0$. Therefore the generators $\{H^\alpha\}$ form a subalgebra of the algebra of the generators of SO(N). The subgroup of the symmetry group generated by them is realized à la Wigner-Weyl.

The remaining generators K^A, with $A = 1, \ldots, \frac{1}{2}N(N-1) - h$, by definition do not preserve the vacuum expectation value of the field

$$(K^A)^i_j\langle\varphi^j\rangle \neq 0. \tag{286}$$

These will be called the *broken generators*. Next we prove a very important result concerning the broken generators known as the Goldstone theorem: for each generator broken by the vacuum expectation value there is a massless excitation.

The mass matrix of the excitations around the vacuum $\langle\varphi^i\rangle$ is determined by the quadratic part of the potential. Since we assumed that $V(\langle\varphi\rangle) = 0$ and we are expanding around a minimum, the first term in the expansion of the potential $V(\varphi)$ around the vacuum expectation values is given by

$$V(\varphi) = \left.\frac{\partial^2 V}{\partial\varphi^i\partial\varphi^j}\right|_{\varphi=\langle\varphi\rangle}(\varphi^i - \langle\varphi^i\rangle)(\varphi^j - \langle\varphi^j\rangle) + \mathcal{O}\left[(\varphi - \langle\varphi\rangle)^3\right] \tag{287}$$

and the mass matrix is:

$$M^2_{ij} \equiv \left.\frac{\partial^2 V}{\partial\varphi^i\partial\varphi^j}\right|_{\varphi=\langle\varphi\rangle}. \tag{288}$$

In order to avoid a cumbersome notation we do not show explicitly the dependence of the mass matrix on the vacuum expectation values $\langle\varphi^i\rangle$.

To extract some information about the possible zero modes of the mass matrix, we write down the conditions that follow from the invariance of the potential under $\delta\varphi^i = \epsilon^a(T^a)^i_j\varphi^j$. At first order in ϵ^a

$$\delta V(\varphi) = \epsilon^a\frac{\partial V}{\partial\varphi^i}(T^a)^i_j\varphi^j = 0. \tag{289}$$

Differentiating this expression with respect to φ^k we arrive at

$$\frac{\partial^2 V}{\partial\varphi^i\partial\varphi^k}(T^a)^i_j\varphi^j + \frac{\partial V}{\partial\varphi^i}(T^a)^i_k = 0. \tag{290}$$

Now we evaluate this expression in the vacuum $\varphi^i = \langle\varphi^i\rangle$. Then the derivative in the second term cancels while the second derivative in the first one gives the mass matrix. Hence we find

$$M^2_{ik}(T^a)^i_j\langle\varphi^j\rangle = 0. \tag{291}$$

Now we can write this expression for both broken and unbroken generators. For the unbroken ones, since $(H^\alpha)^i_j\langle\varphi^j\rangle = 0$, we find a trivial identity $0 = 0$. On the other hand for the broken generators we have

$$M^2_{ik}(K^A)^i_j\langle\varphi^j\rangle = 0. \tag{292}$$

Since $(K^A)^i_j\langle\varphi^j\rangle \neq 0$ this equation implies that the mass matrix has as many zero modes as broken generators. Therefore we have proven Goldstone's theorem: associated with each broken symmetry there is a massless mode in the theory. Here we have presented a classical proof of the theorem. In the quantum theory the proof follows the same lines as the one presented here but one has to consider the effective action containing the effects of the quantum corrections to the classical Lagrangian.

As an example to illustrate this theorem, we consider a SO(3) invariant scalar field theory with a "mexican hat" potential

$$V(\vec{\varphi}) = \frac{\lambda}{4}\left(\vec{\varphi}^{\,2} - a^2\right)^2 . \tag{293}$$

The vacua of the theory correspond to the configurations satisfying $\langle\vec{\varphi}\rangle^{\,2} = a^2$. In field space this equation describes a two-dimensional sphere and each solution is just a point in that sphere. Geometrically it is easy to visualize that a given vacuum field configuration, i.e. a point in the sphere, is preserved by SO(2) rotations around the axis of the sphere that passes through that point. Hence the vacuum expectation value of the scalar field breaks the symmetry according to

$$\langle\vec{\varphi}\rangle : \quad \mathrm{SO}(3) \longrightarrow \mathrm{SO}(2). \tag{294}$$

Since SO(3) has three generators and SO(2) only one we see that two generators are broken and therefore there are two massless Goldstone bosons. Physically this massless modes can be thought of as corresponding to excitations along the surface of the sphere $\langle\vec{\varphi}\rangle^{\,2} = a^2$.

Once a minimum of the potential has been chosen we can proceed to quantize the excitations around it. Since the vacuum only leaves invariant a SO(2) subgroup of the original SO(3) symmetry group it seems that the fact that we are expanding around a particular vacuum expectation value of the scalar field has resulted in a lost of symmetry. This is however not the case. The full quantum theory is symmetric under the whole symmetry group SO(3). This is reflected in the fact that the physical properties of the theory do not depend on the particular point of the sphere $\langle\vec{\varphi}\rangle^{\,2} = a^2$ that we have chosen. Different vacua are related by the full SO(3) symmetry and therefore should give the same physics.

It is very important to realize that given a theory with a vacuum determined by $\langle\vec{\varphi}\rangle$ all other possible vacua of the theory are unaccessible in the infinite volume limit. This means that two vacuum states $|0_1\rangle$, $|0_2\rangle$ corresponding to different vacuum expectation values of the scalar field are orthogonal $\langle 0_1|0_2\rangle = 0$ and cannot be connected by any local observable $\Phi(x)$, $\langle 0_1|\Phi(x)|0_2\rangle = 0$. Heuristically this can be understood by noticing that in the infinite volume limit switching from one vacuum into another one requires changing the vacuum expectation value of the field everywhere in space at the same time, something that cannot be done by any local operator. Notice that this is radically different to our expectations based on the Quantum Mechanics of a system with a finite number of degrees of freedom.

In High Energy Physics the typical example of a Goldstone boson is the pion, associated with the spontaneous breaking of the global chiral isospin $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$ symmetry. This symmetry acts independently in the left- and right-handed spinors as

$$\begin{pmatrix} u_{L,R} \\ d_{L,R} \end{pmatrix} \longrightarrow M_{L,R} \begin{pmatrix} u_{L,R} \\ d_{L,R} \end{pmatrix}, \qquad M_{L,R} \in \mathrm{SU}(2)_{L,R} \tag{295}$$

Presumably since the quarks are confined at low energies this symmetry is spontaneously broken down to the diagonal SU(2) acting in the same way on the left- and right-handed components of the spinors. Associated with this symmetry breaking there is a Goldstone mode which is identified as the pion. Notice, nevertheless, that the $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$ would be an exact global symmetry of the QCD Lagrangian only in the limit when the masses of the quarks are zero $m_u, m_d \to 0$. Since these quarks have nonzero masses the chiral symmetry is only approximate and as a consequence the corresponding Goldstone boson is not massless. That is why pions have masses, although they are the lightest particle among the hadrons.

Symmetry breaking appears also in many places in condensed matter. For example, when a solid crystallizes from a liquid the translational invariance that is present in the liquid phase is broken to a discrete group of translations that represent the crystal lattice. This symmetry breaking has Goldstone

bosons associated which are identified with phonons which are the quantum excitation modes of the vibrational degrees of freedom of the lattice.

The Higgs mechanism. Gauge symmetry seems to prevent a vector field from having a mass. This is obvious once we realize that a term in the Lagrangian like $m^2 A_\mu A^\mu$ is incompatible with gauge invariance.

However certain physical situations seem to require massive vector fields. This happened for example during the 1960s in the study of weak interactions. The Glashow model gave a common description of both electromagnetic and weak interactions based on a gauge theory with group SU(2)×U(1) but, in order to reproduce Fermi's four-fermion theory of the β-decay it was necessary that two of the vector fields involved would be massive. Also in condensed matter physics massive vector fields are required to describe certain systems, most notably in superconductivity.

The way out to this situation is found in the concept of spontaneous symmetry breaking discussed previously. The consistency of the quantum theory requires gauge invariance, but this invariance can be realized à la Nambu-Goldstone. When this is the case the full gauge symmetry is not explicitly present in the effective action constructed around the particular vacuum chosen by the theory. This makes possible the existence of mass terms for gauge fields without jeopardizing the consistency of the full theory, which is still invariant under the whole gauge group.

To illustrate the Higgs mechanism we study the simplest example, the Abelian Higgs model: a U(1) gauge field coupled to a self-interacting charged complex scalar field Φ with Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{D_\mu \Phi} D^\mu \Phi - \frac{\lambda}{4}\left(\overline{\Phi}\Phi - \mu^2\right)^2, \tag{296}$$

where the covariant derivative is given by Eq. (159). This theory is invariant under the gauge transformations

$$\Phi \to e^{i\alpha(x)}\Phi, \qquad A_\mu \to A_\mu + \partial_\mu \alpha(x). \tag{297}$$

The minimum of the potential is defined by the equation $|\Phi| = \mu$. We have a continuum of different vacua labelled by the phase of the scalar field. None of these vacua, however, is invariant under the gauge symmetry

$$\langle \Phi \rangle = \mu e^{i\vartheta_0} \to \mu e^{i\vartheta_0 + i\alpha(x)} \tag{298}$$

and therefore the symmetry is spontaneously broken Let us study now the theory around one of these vacua, for example $\langle \Phi \rangle = \mu$, by writing the field Φ in terms of the excitations around this particular vacuum

$$\Phi(x) = \left[\mu + \frac{1}{\sqrt{2}}\sigma(x)\right] e^{i\vartheta(x)}. \tag{299}$$

Independently of whether we are expanding around a particular vacuum for the scalar field we should keep in mind that the whole Lagrangian is still gauge invariant under (297). This means that performing a gauge transformation with parameter $\alpha(x) = -\vartheta(x)$ we can get rid of the phase in Eq. (299). Substituting then $\Phi(x) = \mu + \frac{1}{\sqrt{2}}\sigma(x)$ in the Lagrangian we find

$$\begin{aligned} \mathcal{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + e^2\mu^2 A_\mu A^\mu + \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma - \frac{1}{2}\lambda\mu^2\sigma^2 \\ &- \lambda\mu\sigma^3 - \frac{\lambda}{4}\sigma^4 + e^2\mu A_\mu A^\mu \sigma + e^2 A_\mu A^\mu \sigma^2. \end{aligned} \tag{300}$$

What are the excitation of the theory around the vacuum $\langle \Phi \rangle = \mu$? First we find a massive real scalar field $\sigma(x)$. The important point however is that the vector field A_μ now has a mass given by

$$m_\gamma^2 = 2e^2\mu^2. \tag{301}$$

The remarkable thing about this way of giving a mass to the photon is that at no point we have given up gauge invariance. The symmetry is only hidden. Therefore in quantizing the theory we can still enjoy all the advantages of having a gauge theory but at the same time we have managed to generate a mass for the gauge field.

It is surprising, however, that in the Lagrangian (300) we did not found any massless mode. Since the vacuum chosen by the scalar field breaks the $U(1)$ generator of U(1) we would have expected one masless particle from Goldstone's theorem. To understand the fate of the missing Goldstone boson we have to revisit the calculation leading to Eq. (300). Were we dealing with a global U(1) theory, the Goldstone boson would correspond to excitation of the scalar field along the valley of the potential and the phase $\vartheta(x)$ would be the massless Goldstone boson. However we have to keep in mind that in computing the Lagrangian we managed to get rid of $\vartheta(x)$ by shifting it into A_μ using a gauge transformation. Actually by identifying the gauge parameter with the Goldstone excitation we have completely fixed the gauge and the Lagrangian (300) does not have any gauge symmetry left.

A massive vector field has three polarizations: two transverse ones $\vec{k}\cdot\vec{\epsilon}(\vec{k},\pm 1)=0$ plus a longitudinal one $\vec{\epsilon}_L(\vec{k})\sim\vec{k}$. In gauging away the massless Goldstone boson $\vartheta(x)$ we have transformed it into the longitudinal polarization of the massive vector field. In the literature this is usually expressed saying that the Goldstone mode is "eaten up" by the longitudinal component of the gauge field. It is important to realize that in spite of the fact that the Lagrangian (300) looks pretty different from the one we started with we have not lost any degrees of freedom. We started with the two polarizations of the photon plus the two degrees of freedom associated with the real and imaginary components of the complex scalar field. After symmetry breaking we end up with the three polarizations of the massive vector field and the degree of freedom of the real scalar field $\sigma(x)$.

We can also understand the Higgs mechanism in the light of our discussion of gauge symmetry in section 4.4. In the Higgs mechanism the invariance of the theory under infinitesimal gauge transformations is not explicitly broken, and this implies that Gauss' law is satisfied quantum mechanically, $\vec{\nabla}\cdot\vec{E}_a|\text{phys}\rangle=0$. The theory remains invariant under gauge transformations in the connected component of the identity $\mathcal{G}_0$, the ones generated by Gauss' law. This does not pose any restriction on the possible breaking of the invariance of the theory with respect to transformations that cannot be continuously deformed to the identity. Hence in the Higgs mechanism the invariance under gauge transformation that are not in the connected component of the identity, $\mathcal{G}/\mathcal{G}_0$, can be broken. Let us try to put it in more precise terms. As we learned in section 4.4, in the Hamiltonian formulation of the theory finite energy gauge field configurations tend to a pure gauge at spatial infinity

$$\vec{A}_\mu(\vec{x})\longrightarrow -\frac{1}{ig}g(\vec{x})\vec{\nabla}g(\vec{x})^{-1},\qquad |\vec{x}|\to\infty \tag{302}$$

The set transformations $g_0(\vec{x})\in\mathcal{G}_0$ that tend to the identity at infinity are the ones generated by Gauss' law. However, one can also consider in general gauge transformations $g(\vec{x})$ which, as $|\vec{x}|\to\infty$, approach any other element $g\in G$. The quotient $\mathcal{G}_\infty\equiv\mathcal{G}/\mathcal{G}_0$ gives a copy of the gauge group at infinity. There is no reason, however, why this group should not be broken, and in general it is if the gauge symmetry is spontaneously broken. Notice that this is not a threat to the consistency of the theory. Properties like the decoupling of unphysical states are guaranteed by the fact that Gauss' law is satisfied quantum mechanically and are not affected by the breaking of $\mathcal{G}_\infty$.

In condensed matter physics the symmetry breaking described by the nonrelativistic version of the Abelian Higgs model can be used to characterize the onset of a superconducting phase in the BCS theory, where the complex scalar field Φ is associated with the Cooper pairs. In this case the parameter μ^2 depends on the temperature. Above the critical temperature T_c, $\mu^2(T)>0$ and there is only a symmetric vacuum $\langle\Phi\rangle=0$. When, on the other hand, $T<T_c$ then $\mu^2(T)<0$ and symmetry breaking takes place. The onset of a nonzero mass of the photon (301) below the critical temperature explains the Meissner effect: the magnetic fields cannot penetrate inside superconductors beyond a distance of the order $\frac{1}{m_\gamma}$.

The Abelian Higgs model discussed here can be regarded as a toy model of the Brout-Englert-Higgs mechanism responsible for giving mass to the $W^{\pm}$ and Z^0 gauge bosons in the standard model. Giving mass to these three bosons requires the introduction of a two-component complex scalar field transforming as a doublet under SU(2). Three of its four degrees of freedom are incorporated as the longitudinal components of the three massive gauge fields, whereas the fourth one remains as a scalar propagating degree of freedom. Its elementary excitations are spin zero neutral particles known as Higgs bosons.

The Higgs boson couples to the massive gauge fields, as well as to quarks and leptons. Moreover, its coupling to the fermions is proportional to the fermion masses and therefore very weak for light fermions. This, together with the fact that Higgs productions processes have large standard model backgrounds, complicates its experimental detection. After decades of searches in various experiments, a Higgs boson candidate was finally detected at the ATLAS and CMS collaborations at the Large Hadron Collider (LHC) in 2012 with a mass of approximately 125 GeV. At the time of writing, all evidences point to the fact that this new particle is indeed the so much coveted standard model Higgs.

7 Anomalies

So far we did not worry too much about how classical symmetries of a theory are carried over to the quantum theory. We have implicitly assumed that classical symmetries are preserved in the process of quantization, so they are also realized in the quantum theory.

This, however, does not have to be necessarily the case. Quantizing an interacting field theory is a very involved process that requires regularization and renormalization and sometimes, it does not matter how hard we try, there is no way for a classical symmetry to survive quantization. When this happens one says that the theory has an *anomaly* (for reviews see [28]). It is important to avoid here the misconception that anomalies appear due to a bad choice of the way a theory is regularized in the process of quantization. When we talk about anomalies we mean a classical symmetry that *cannot* be realized in the quantum theory, no matter how smart we are in choosing the regularization procedure.

In the following we analyze some examples of anomalies associated with global and local symmetries of the classical theory. In Section 8 we will encounter yet another example of an anomaly, this time associated with the breaking of classical scale invariance in the quantum theory.

7.1 Axial anomaly

Probably the best known examples of anomalies appear when we consider axial symmetries. If we consider a theory of two Weyl spinors $u_{\pm}$

$$\mathcal{L} = i\overline{\psi}\partial\!\!\!/\psi = iu_+^\dagger\sigma_+^\mu\partial_\mu u_+ + iu_-^\dagger\sigma_-^\mu\partial_\mu u_- \qquad \text{with} \qquad \psi = \begin{pmatrix} u_+ \\ u_- \end{pmatrix} \tag{303}$$

the Lagrangian is invariant under two types of global U(1) transformations. In the first one both helicities transform with the same phase, this is a *vector* transformation:

$$\text{U(1)}_V : u_{\pm} \longrightarrow e^{i\alpha}u_{\pm}, \tag{304}$$

whereas in the second one, the axial $U(1)$, the signs of the phases are different for the two chiralities

$$\text{U(1)}_A : u_{\pm} \longrightarrow e^{\pm i\alpha}u_{\pm}. \tag{305}$$

Using Noether's theorem, there are two conserved currents, a vector current

$$J_V^\mu = \overline{\psi}\gamma^\mu\psi = u_+^\dagger\sigma_+^\mu u_+ + u_-^\dagger\sigma_-^\mu u_- \qquad \Longrightarrow \qquad \partial_\mu J_V^\mu = 0 \tag{306}$$

and an axial vector current

$$J_A^\mu = \overline{\psi}\gamma^\mu\gamma_5\psi = u_+^\dagger\sigma_+^\mu u_+ - u_-^\dagger\sigma_-^\mu u_- \qquad \Longrightarrow \qquad \partial_\mu J_A^\mu = 0. \tag{307}$$

The theory described by the Lagrangian (303) can be coupled to the electromagnetic field. The resulting classical theory is still invariant under the vector and axial U(1) symmetries (304) and (305). Surprisingly, upon quantization it turns out that the conservation of the axial current (307) is spoiled by quantum effects

$$\partial_\mu J_A^\mu \sim \hbar\,\vec{E}\cdot\vec{B}. \tag{308}$$

To understand more clearly how this result comes about we study first a simple model in two dimensions that captures the relevant physics involved in the four-dimensional case [29]. We work in Minkowski space in two dimensions with coordinates $(x^0, x^1) \equiv (t, x)$ and where the spatial direction is compactified to a circle S^1. In this setup we consider a fermion coupled to the electromagnetic field. Notice that since we are living in two dimensions the field strength $F_{\mu\nu}$ only has one independent component that corresponds to the electric field along the spatial direction, $F^{01} \equiv \mathcal{E}$ (in two dimensions there are no magnetic fields!).

To write the Lagrangian for the spinor field we need to find a representation of the algebra of γ-matrices

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \quad \text{with} \quad \eta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{309}$$

In two dimensions the dimension of the representation of the γ-matrices is $2^{[\frac{2}{2}]} = 2$. Here take

$$\gamma^0 \equiv \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \gamma^1 \equiv i\sigma^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{310}$$

This is a chiral representation since the matrix γ_5 is diagonal[15]

$$\gamma_5 \equiv -\gamma^0\gamma^1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{311}$$

Writing the two-component spinor ψ as

$$\psi = \begin{pmatrix} u_+ \\ u_- \end{pmatrix} \tag{312}$$

and defining as usual the projectors $P_\pm = \frac{1}{2}(\mathbf{1} \pm \gamma_5)$ we find that the components $u_\pm$ of ψ are respectively a right- and left-handed Weyl spinor in two dimensions.

Once we have a representation of the γ-matrices we can write the Dirac equation. Expressing it in terms of the components $u_\pm$ of the Dirac spinor we find

$$(\partial_0 - \partial_1)u_+ = 0, \qquad (\partial_0 + \partial_1)u_- = 0. \tag{313}$$

The general solution to these equations can be immediately written as

$$u_+ = u_+(x^0 + x^1), \qquad u_- = u_-(x^0 - x^1). \tag{314}$$

Hence $u_\pm$ are two wave packets moving along the spatial dimension respectively to the left (u_+) and to the right (u_-). Notice that according to our convention the left-moving u_+ is a right-handed spinor (positive helicity) whereas the right-moving u_- is a left-handed spinor (negative helicity).

[15] In any even number of dimensions γ_5 is defined to satisfy the conditions $\gamma_5^2 = \mathbf{1}$ and $\{\gamma_5, \gamma^\mu\} = 0$.

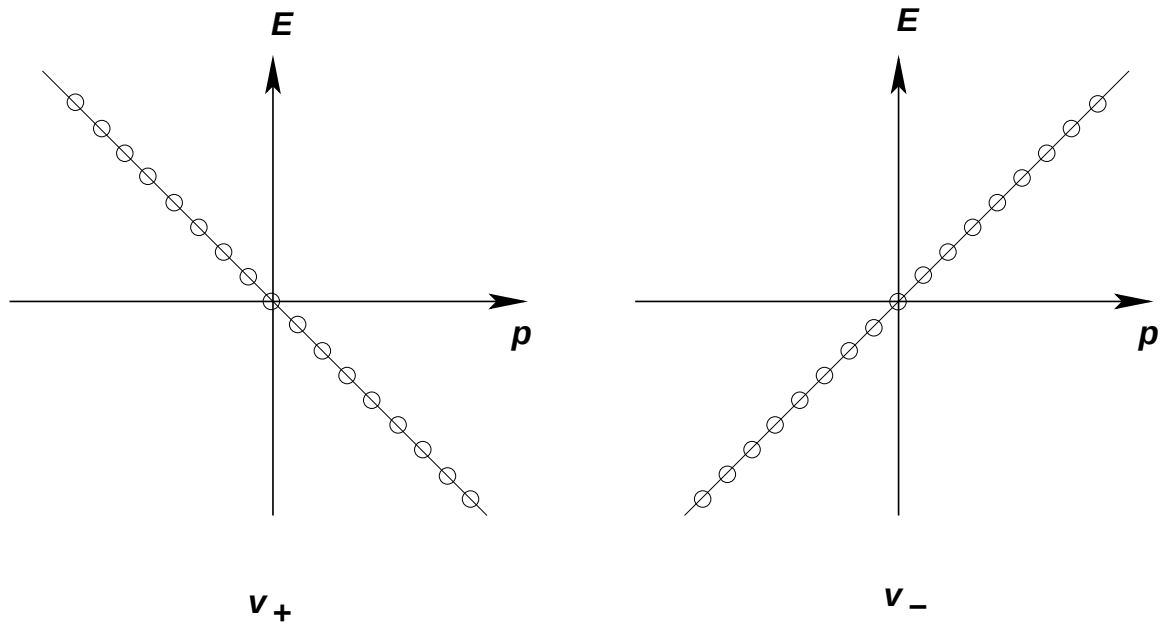


Fig. 11: Spectrum of the massless two-dimensional Dirac field.

If we want to interpret (313) as the wave equation for two-dimensional Weyl spinors we have the following wave functions for free particles with well defined momentum $p^\mu = (E, p)$.

$$u_\pm^{(E)}(x^0 \pm x^1) = \frac{1}{\sqrt{L}} e^{-iE(x^0 \pm x^1)} \quad \text{with} \quad p = \mp E. \tag{315}$$

As it is always the case with the Dirac equation we have both positive and negative energy solutions. For u_+, since $E = -p$, we see that the solutions with positive energy are those with negative momentum $p < 0$, whereas the negative energy solutions are plane waves with $p > 0$. For the left-handed spinor u_- the situation is reversed. Besides, since the spatial direction is compact with length L the momentum p is quantized according to

$$p = \frac{2\pi n}{L}, \qquad n \in \mathbb{Z}. \tag{316}$$

The spectrum of the theory is represented in Fig. 11.

Once we have the spectrum of the theory the next step is to obtain the vacuum. As with the Dirac equation in four dimensions we fill all the states with $E \leq 0$ (Fig. 12). Exciting of a particle in the Dirac see produces a positive energy fermion plus a hole that is interpreted as an antiparticle. This gives us the clue on how to quantize the theory. In the expansion of the operator $u_\pm$ in terms of the modes (315) we associate positive energy states with annihilation operators whereas the states with negative energy are associated with creation operators for the corresponding antiparticle

$$u_\pm(x) = \sum_{E>0} \left[a_\pm(E) v_\pm^{(E)}(x) + b_\pm^\dagger(E) v_\pm^{(E)}(x)^* \right]. \tag{317}$$

The operator $a_\pm(E)$ acting on the vacuum $|0, \pm\rangle$ annihilates a particle with positive energy E and momentum $\mp E$. In the same way $b_\pm^\dagger(E)$ creates out of the vacuum an antiparticle with positive energy E and spatial momentum $\mp E$. In the Dirac sea picture the operator $b_\pm(E)^\dagger$ is originally an annihilation operator for a state of the sea with negative energy $-E$. As in the four-dimensional case the problem of the negative energy states is solved by interpreting annihilation operators for negative energy states as

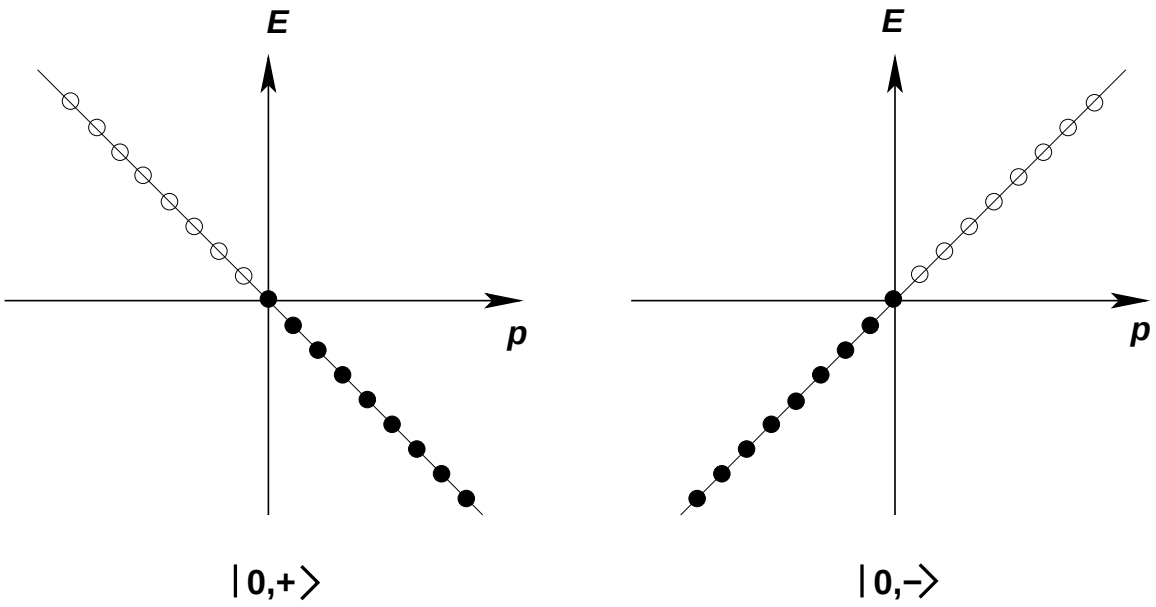


Fig. 12: Vacuum of the theory.

creation operators for the corresponding antiparticle with positive energy (and vice versa). The operators appearing in the expansion of $u_\pm$ in Eq. (317) satisfy the usual algebra

$$\{a_\lambda(E), a^\dagger_{\lambda'}(E')\} = \{b_\lambda(E), b^\dagger_{\lambda'}(E')\} = \delta_{E,E'}\delta_{\lambda\lambda'}, \tag{318}$$

where we have introduced the label $\lambda, \lambda' = \pm$. Also, $a_\lambda(E)$, $a^\dagger_\lambda(E)$ anticommute with $b_{\lambda'}(E')$, $b^\dagger_{\lambda'}(E')$.

The Lagrangian of the theory

$$\mathcal{L} = iu^\dagger_+(\partial_0 + \partial_1)u_+ + iu^\dagger_-(\partial_0 - \partial_1)u_- \tag{319}$$

is invariant under both U(1)$_V$, Eq. (304), and U(1)$_A$, Eq. (305). The associated Noether currents are in this case

$$J^\mu_V = \begin{pmatrix} u^\dagger_+u_+ + u^\dagger_-u_- \\ -u^\dagger_+u_+ + u^\dagger_-u_- \end{pmatrix}, \qquad J^\mu_A = \begin{pmatrix} u^\dagger_+u_+ - u^\dagger_-u_- \\ -u^\dagger_+u_+ - u^\dagger_-u_- \end{pmatrix}. \tag{320}$$

The associated conserved charges are given, for the vector current by

$$Q_V = \int_0^L dx^1 \left(u^\dagger_+u_+ + u^\dagger_-u_-\right) \tag{321}$$

and for the axial current

$$Q_A = \int_0^L dx^1 \left(u^\dagger_+u_+ - u^\dagger_-u_-\right). \tag{322}$$

Using the orthonormality relations for the modes $v^{(E)}_\pm(x)$

$$\int_0^L dx^1\, v^{(E)}_\pm(x)\, v^{(E')}_\pm(x) = \delta_{E,E'} \tag{323}$$

we find for the conserved charges:

$$Q_V \;=\; \sum_{E>0}\left[a^\dagger_+(E)a_+(E) - b^\dagger_+(E)b_+(E) + a^\dagger_-(E)a_-(E) - b^\dagger_-(E)b_-(E)\right],$$

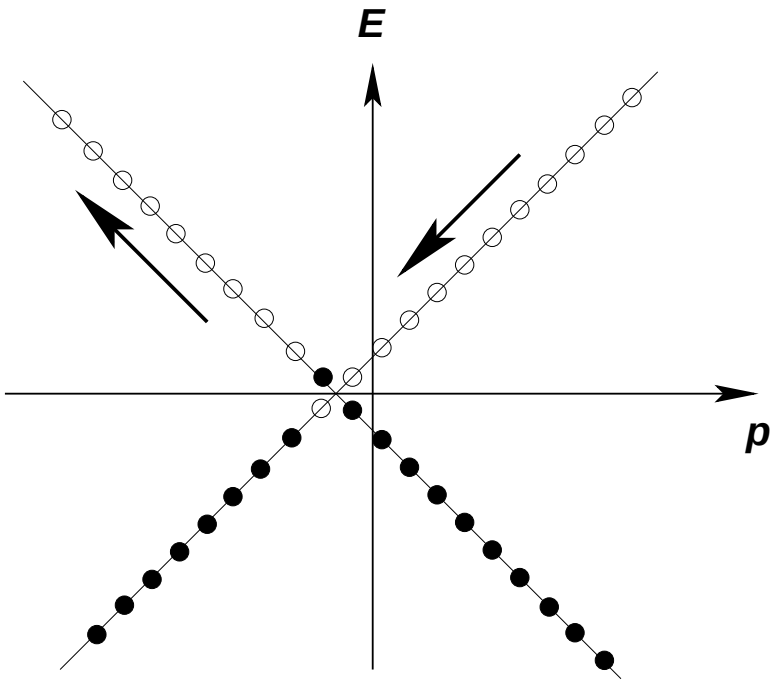


Fig. 13: Effect of the electric field.

$$Q_A \;=\; \sum_{E>0}\Big[a_+^\dagger(E)a_+(E) - b_+^\dagger(E)b_+(E) - a_-^\dagger(E)a_-(E) + b_-^\dagger(E)b_-(E)\Big]\,. \qquad (324)$$

We see that Q_V counts the net number (particles minus antiparticles) of positive helicity states plus the net number of states with negative helicity. The axial charge, on the other hand, counts the net number of positive helicity states minus the number of negative helicity ones. In the case of the vector current we have subtracted a formally divergent vacuum contribution to the charge (the "charge of the Dirac sea").

In the free theory there is of course no problem with the conservation of either Q_V or Q_A, since the occupation numbers do not change. What we want to study is the effect of coupling the theory to electric field $\mathcal{E}$. We work in the gauge $A_0 = 0$. Instead of solving the problem exactly we are going to simulate the electric field by adiabatically varying in a long time τ_0 the vector potential A_1 from zero value to $-\mathcal{E}\tau_0$. From our discussion in section 4.3 we know that the effect of the electromagnetic coupling in the theory is a shift in the momentum according to

$$p \longrightarrow p - eA_1, \qquad (325)$$

where e is the charge of the fermions. Since we assumed that the vector potential varies adiabatically, we can assume it to be approximately constant at each time.

Then, we have to understand what is the effect of (325) on the vacuum depicted in Fig. (12). What we find is that the two branches move as shown in Fig. (13) resulting in some of the negative energy states of the v_+ branch acquiring positive energy while the same number of the empty positive energy states of the other branch v_- will become empty negative energy states. Physically this means that the external electric field $\mathcal{E}$ creates a number of particle-antiparticle pairs out of the vacuum. Denoting by $N \sim e\mathcal{E}$ the number of such pairs created by the electric field per unit time, the final values of the charges Q_V and Q_A are

$$\begin{aligned} Q_A(\tau_0) &= (N-0) + (0-N) = 0, \\ Q_V(\tau_0) &= (N-0) - (0-N) = 2N. \end{aligned} \qquad (326)$$

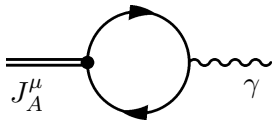
Therefore we conclude that the coupling to the electric field produces a violation in the conservation of the axial charge per unit time given by $\Delta Q_A \sim e\mathcal{E}$. This implies that

$$\partial_\mu J^\mu_A \sim e\hbar\mathcal{E}, \tag{327}$$

where we have restored $\hbar$ to make clear that the violation in the conservation of the axial current is a quantum effect. At the same time $\Delta Q_V = 0$ guarantees that the vector current remains conserved also quantum mechanically, $\partial_\mu J^\mu_V = 0$.

We have just studied a two-dimensional example of the Adler-Bell-Jackiw axial anomaly [30]. The heuristic analysis presented here can be made more precise by computing the quantity

$$C^{\mu\nu} = \langle 0|T\left[J^\mu_A(x)J^\nu_V(0)\right]|0\rangle = \tag{328}$$



The anomaly is given then by $\partial_\mu C^{\mu\nu}$. A careful calculation yields the numerical prefactor missing in Eq. (327) leading to the result

$$\partial_\mu J^\mu_A = \frac{e\hbar}{2\pi}\varepsilon^{\nu\sigma}F_{\nu\sigma}, \tag{329}$$

with $\varepsilon^{01} = -\varepsilon^{10} = 1$.

The existence of an anomaly in the axial symmetry that we have illustrated in two dimensions is present in all even dimensional of space-times. In particular in four dimensions the axial anomaly it is given by

$$\partial_\mu J^\mu_A = -\frac{e^2}{16\pi^2}\varepsilon^{\mu\nu\sigma\lambda}F_{\mu\nu}F_{\sigma\lambda}. \tag{330}$$

This result has very important consequences in the physics of strong interactions as we will see in what follows

7.2 Chiral symmetry in QCD

Our knowledge of the physics of strong interactions is based on the theory of Quantum Chromodynamics (QCD) [32]. This is a nonabelian gauge theory with gauge group SU(N_c) coupled to a number N_f of quarks. These are spin-$\frac{1}{2}$ particles $Q^{i\,f}$ labelled by two quantum numbers: color $i = 1, \ldots, N_c$ and flavor $f = 1, \ldots, N_f$. The interaction between them is mediated by the $N_c^2 - 1$ gauge bosons, the gluons A^a_μ, $a = 1, \ldots, N_c^2 - 1$. In the real world $N_c = 3$ and the number of flavors is six, corresponding to the number of different quarks: up (u), down (d), charm (c), strange (s), top (t) and bottom (b).

For the time being we are going to study a general theory of QCD with N_c colors and N_f flavors. Also, for reasons that will be clear later we are going to work in the limit of vanishing quark masses, $m_f \to 0$. In this cases the Lagrangian is given by

$$\mathcal{L}_{\rm QCD} = -\frac{1}{4}F^a_{\mu\nu}F^{a\,\mu\nu} + \sum_{f=1}^{N_f}\left[i\overline{Q}^f_L \not{D}\, Q^f_L + i\overline{Q}^f_R \not{D}\, Q^f_R\right], \tag{331}$$

where the subscripts L and R indicate respectively left and right-handed spinors, $Q^f_{L,R} \equiv P_\pm Q^f$, and the field strength $F^a_{\mu\nu}$ and the covariant derivative D_μ are respectively defined in Eqs. (165) and (168). Apart

from the gauge symmetry, this Lagrangian is also invariant under a global $\mathrm{U}(N_f)_L \times \mathrm{U}(N_f)_R$ acting on the flavor indices and defined by

$$\mathrm{U}(N_f)_L : \begin{cases} Q_L^f & \to & \sum_{f'} (U_L)_{ff'} Q_L^{f'} \\ Q_R^f & \to & Q_R^f \end{cases} \qquad \mathrm{U}(N_f)_R : \begin{cases} Q_L^f & \to & Q_L^f \\ Q_R^r & \to & \sum_{f'} (U_R)_{ff'} Q_R^{f'} \end{cases} \tag{332}$$

with $U_L, U_R \in \mathrm{U}(N_f)$. Actually, since $\mathrm{U}(N)=\mathrm{U}(1)\times\mathrm{SU}(N)$ this global symmetry group can be written as $\mathrm{SU}(N_f)_L \times \mathrm{SU}(N_f)_R \times \mathrm{U}(1)_L \times \mathrm{U}(1)_R$. The abelian subgroup $\mathrm{U}(1)_L \times \mathrm{U}(1)_R$ can be now decomposed into their vector $\mathrm{U}(1)_B$ and axial $\mathrm{U}(1)_A$ subgroups defined by the transformations

$$\mathrm{U}(1)_B : \begin{cases} Q_L^f & \to & e^{i\alpha} Q_L^f \\ Q_R^f & \to & e^{i\alpha} Q_R^f \end{cases} \qquad \mathrm{U}(1)_A : \begin{cases} Q_L^f & \to & e^{i\alpha} Q_L^f \\ Q_R^f & \to & e^{-i\alpha} Q_R^f \end{cases} \tag{333}$$

According to Noether's theorem, associated with these two abelian symmetries we have two conserved currents:

$$J_V^\mu = \sum_{f=1}^{N_f} \overline{Q}^f \gamma^\mu \, Q^f, \qquad J_A^\mu = \sum_{f=1}^{N_f} \overline{Q}^f \gamma^\mu \gamma_5 \, Q^f. \tag{334}$$

The conserved charge associated with vector charge J_V^μ is actually the baryon number defined as the number of quarks minus number of antiquarks.

The nonabelian part of the global symmetry group $\mathrm{SU}(N_f)_L \times \mathrm{SU}(N_f)_R$ can also be decomposed into its vector and axial subgroups, $\mathrm{SU}(N_f)_V \times \mathrm{SU}(N_f)_A$, defined by the following transformations of the quarks fields

$$\mathrm{SU}(N_f)_V : \begin{cases} Q_L^f & \to & \sum_{f'} (U_L)_{ff'} Q_L^{f'} \\ Q_R^f & \to & \sum_{f'} (U_L)_{ff'} Q_R^{f'} \end{cases} \qquad \mathrm{SU}(N_f)_A : \begin{cases} Q_L^f & \to & \sum_{f'} (U_L)_{ff'} Q_L^{f'} \\ Q_R^f & \to & \sum_{f'} (U_R^{-1})_{ff'} Q_R^{f'} \end{cases} \tag{335}$$

Again, the application of Noether's theorem shows the existence of the following nonabelian conserved charges

$$J_V^{I\,\mu} \equiv \sum_{f,f'=1}^{N_f} \overline{Q}^f \gamma^\mu (T^I)_{ff'} Q^{f'}, \qquad J_A^{I\,\mu} \equiv \sum_{f,f'=1}^{N_f} \overline{Q}^f \gamma^\mu \gamma_5 (T^I)_{ff'} Q^{f'}. \tag{336}$$

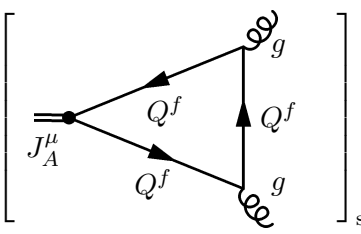
To summarize, we have shown that the initial chiral symmetry of the QCD Lagrangian (331) can be decomposed into its chiral and vector subgroups according to

$$\mathrm{U}(N_f)_L \times \mathrm{U}(N_f)_R = \mathrm{SU}(N_f)_V \times \mathrm{SU}(N_f)_A \times \mathrm{U}(1)_B \times \mathrm{U}(1)_A. \tag{337}$$

The question to address now is which part of the classical global symmetry is preserved by the quantum theory.

As argued in section 7.1, the conservation of the axial currents J_A^μ and $J_A^{a\,\mu}$ can in principle be spoiled due to the presence of an anomaly. In the case of the abelian axial current J_A^μ the relevant quantity is the correlation function

$$C^{\mu\nu\sigma} \equiv \langle 0 | T \left[J_A^\mu(x) j_{\mathrm{gauge}}^{a\,\nu}(x') j_{\mathrm{gauge}}^{b\,\sigma}(0) \right] |0\rangle = \sum_{f=1}^{N_f} \Bigg[\quad \Bigg]_{\mathrm{symmetric}} \tag{338}$$



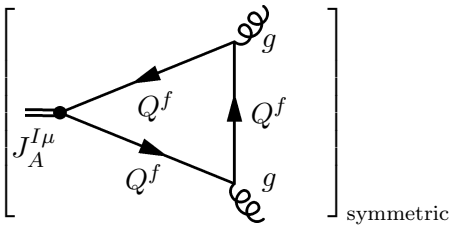
Here $j^{a\,\mu}_{\rm gauge}$ is the nonabelian conserved current coupling to the gluon field

$$j^{a\,\mu}_{\rm gauge} \equiv \sum_{f=1}^{N_f} \overline{Q}^f \gamma^\mu \tau^a Q^f, \tag{339}$$

where, to avoid confusion with the generators of the global symmetry we have denoted by τ^a the generators of the gauge group SU(N_c). The anomaly can be read now from $\partial_\mu C^{\mu\nu\sigma}$. If we impose Bose symmetry with respect to the interchange of the two outgoing gluons and gauge invariance of the whole expression, $\partial_\nu C^{\mu\nu\sigma} = 0 = \partial_\sigma C^{\mu\nu\sigma}$, we find that the axial abelian global current has an anomaly given by[16]

$$\partial_\mu J^\mu_A = -\frac{g^2 N_f}{32\pi^2} \varepsilon^{\mu\nu\sigma\lambda} F^a_{\mu\nu} F^{a\,\mu\nu}. \tag{340}$$

In the case of the nonabelian axial global symmetry SU(N_f)$_A$ the calculation of the anomaly is made as above. The result, however, is quite different since in this case we conclude that the nonabelian axial current $J^{a\,\mu}_A$ is not anomalous. This can be easily seen by noticing that associated with the axial current vertex we have a generator T^I of SU(N_f), whereas for the two gluon vertices we have the generators τ^a of the gauge group SU(N_c). Therefore, the triangle diagram is proportional to the group-theoretic factor

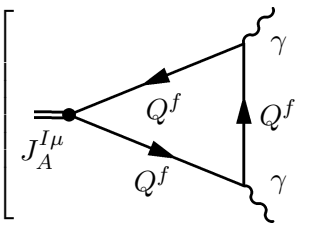


$$\sim {\rm tr}\, T^I \ {\rm tr}\, \{\tau^a, \tau^b\} = 0 \tag{341}$$

which vanishes because the generators of SU(N_f) are traceless.

From here we would conclude that the nonabelian axial symmetry SU(N_f)$_A$ is nonanomalous. However this is not the whole story since quarks are charged particles that also couple to photons. Hence there is a second potential source of an anomaly coming from the the one-loop triangle diagram coupling $J^{I\,\mu}_A$ to two photons

$$\langle 0|T\left[J^{I\,\mu}_A(x) j^\nu_{\rm em}(x') j^\sigma_{\rm em}(0)\right]|0\rangle = \sum_{f=1}^{N_f} \left[\quad \right]_{\rm symmetric} \tag{342}$$



where $j^\mu_{\rm em}$ is the electromagnetic current

$$j^\mu_{\rm em} = \sum_{f=1}^{N_f} q_f\, \overline{Q}^f \gamma^\mu Q^f, \tag{343}$$

with q_f the electric charge of the f-th quark flavor. A calculation of the diagram in (342) shows the existence of an Adler-Bell-Jackiw anomaly given by

$$\partial_\mu J^{I\,\mu}_A = -\frac{N_c}{16\pi^2} \left[\sum_{f=1}^{N_f} (T^I)_{ff}\, q_f^2\right] \varepsilon^{\mu\nu\sigma\lambda} F_{\mu\nu} F_{\sigma\lambda}, \tag{344}$$

[16]The normalization of the generators T^I of the global SU(N_f) is given by ${\rm tr}\,(T^I T^J) = \frac{1}{2}\delta^{IJ}$.

where $F_{\mu\nu}$ is the field strength of the electromagnetic field coupling to the quarks. The only chance for the anomaly to cancel is that the factor between brackets in this equation be identically zero.

Before proceeding let us summarize the results found so far. Because of the presence of anomalies the axial part of the global chiral symmetry, SU$(N_f)_A$ and U$(1)_A$ are not realized quantum mechanically in general. We found that U$(1)_A$ is always affected by an anomaly. However, because the right-hand side of the anomaly equation (340) is a total derivative, the anomalous character of J^μ_A does not explain the absence of U$(1)_A$ multiplets in the hadron spectrum, since a new current can be constructed which is conserved. In addition, the nonexistence of candidates for a Goldstone boson associated with the right quantum numbers indicates that U$(1)_A$ is not spontaneously broken either, so it has be explicitly broken somehow. This is the so-called U(1)-problem which was solved by 't Hooft [33], who showed how the contribution of quantum transitions between vacua with topologically nontrivial gauge field configurations (instantons) results in an explicit breaking of this symmetry.

Due to the dynamics of the SU(N_c) gauge theory the axial nonabelian symmetry is spontaneously broken due to the presence at low energies of a vacuum expectation value for the fermion bilinear $\overline{Q}^f Q^f$

$$\langle 0|\overline{Q}^f Q^f|0\rangle \neq 0 \qquad \text{(No summation in } f\text{!)}. \tag{345}$$

This nonvanishing vacuum expectation value for the quark bilinear actually breaks chiral invariance spontaneously to the vector subgroup SU$(N_f)_V$, so the only subgroup of the original global symmetry that is realized by the full theory at low energy is

$$\text{U}(N_f)_L \times \text{U}(N_f)_R \longrightarrow \text{SU}(N_f)_V \times \text{U}(1)_B. \tag{346}$$

Associated with this breaking a Goldstone boson should appear with the quantum numbers of the broken nonabelian current. For example, in the case of QCD the Goldstone bosons associated with the spontaneously symmetry breaking induced by the vacuum expectation values $\langle\overline{u}u\rangle$, $\langle\overline{d}d\rangle$ and $\langle(\overline{u}d-\overline{d}u)\rangle$ have been identified as the pions π^0, $\pi^\pm$. These bosons are not exactly massless because of the nonvanishing mass of the u and d quarks. Since the global chiral symmetry is already slightly broken by mass terms in the Lagrangian, the associated Goldstone bosons also have masses although they are very light compared to the masses of other hadrons.

In order to have a better physical understanding of the role of anomalies in the physics of strong interactions we particularize now our analysis of the case of real QCD. Since the u and d quarks are much lighter than the other four flavors, QCD at low energies can be well described by including only these two flavors and ignoring heavier quarks. In this approximation, from our previous discussion we know that the low energy global symmetry of the theory is SU$(2)_V\times$U$(1)_B$, where now the vector group SU$(2)_V$ is the well-known isospin symmetry. The axial U$(1)_A$ current is anomalous due to Eq. (340) with $N_f = 2$. In the case of the nonabelian axial symmetry SU$(2)_A$, taking into account that $q_u = \frac{2}{3}e$ and $q_d = -\frac{1}{3}e$ and that the three generators of SU(2) can be written in terms of the Pauli matrices as $T^K = \frac{1}{2}\sigma^K$ we find

$$\sum_{f=u,d}(T^1)_{ff}\,q_f^2 = \sum_{f=u,d}(T^1)_{ff}\,q_f^2 = 0, \qquad \sum_{f=u,d}(T^3)_{ff}\,q_f^2 = \frac{e^2}{6}. \tag{347}$$

Therefore $J^{3\,\mu}_A$ is anomalous.

Physically, the anomaly in the axial current $J^{3\,\mu}_A$ has an important consequence. In the quark model, the wave function of the neutral pion π^0 is given in terms of those for the u and d quark by

$$|\pi^0\rangle = \frac{1}{\sqrt{2}}\left(|\bar{u}\rangle|u\rangle - |\bar{d}\rangle|d\rangle\right). \tag{348}$$

The isospin quantum numbers of $|\pi^0\rangle$ are those of the generator T^3. Actually the analogy goes further since $\partial_\mu J_A^{3\,\mu}$ is the operator creating a pion π^0 out of the vacuum

$$|\pi^0\rangle \sim \partial_\mu J_A^{3\,\mu}|0\rangle. \tag{349}$$

This leads to the physical interpretation of the triangle diagram (342) with $J_A^{3\,\mu}$ as the one loop contribution to the decay of a neutral pion into two photons

$$\pi^0 \longrightarrow 2\gamma\,. \tag{350}$$

This is an interesting piece of physics. In 1967 Sutherland and Veltman [34] presented a calculation, using current algebra techniques, according to which the decay of the pion into two photons should be suppressed. This however contradicted the experimental evidence that showed the existence of such a decay. The way out to this paradox, as pointed out in [30], is the axial anomaly. What happens is that the current algebra analysis overlooks the ambiguities associated with the regularization of divergences in quantum field theory. A QED evaluation of the triangle diagram leads to a divergent integral that has to be regularized somehow. It is in this process that the Adler-Bell-Jackiw axial anomaly appears resulting in a nonvanishing value for the $\pi^0 \to 2\gamma$ amplitude[17].

The existence of anomalies associated with global currents does not necessarily mean difficulties for the theory. On the contrary, as we saw in the case of the axial anomaly it is its existence what allows for a solution of the Sutherland-Veltman paradox and an explanation of the electromagnetic decay of the pion. The situation, however, is very different if we deal with local symmetries. A quantum mechanical violation of gauge symmetry leads to all kinds of problems, from lack of renormalizability to nondecoupling of negative norm states. This is because the presence of an anomaly in the theory implies that the Gauss' law constraint $\vec{\nabla}\cdot\vec{E}_a = \rho_a$ cannot be consistently implemented in the quantum theory. As a consequence states that classically are eliminated by the gauge symmetry become propagating fields in the quantum theory, thus spoiling the consistency of the theory.

Anomalies in a gauge symmetry can be expected only in chiral theories where left and right-handed fermions transform in different representations of the gauge group. Physically, the most interesting example of such theories is the electroweak sector of the standard model where, for example, left handed fermions transform as doublets under SU(2) whereas right-handed fermions are singlets. On the other hand, QCD is free of gauge anomalies since both left- and right-handed quarks transform in the fundamental representation of SU(3).

We consider the Lagrangian

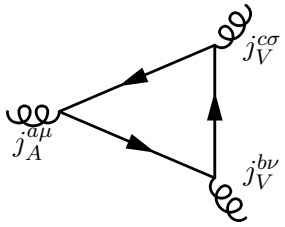
$$\mathcal{L} = -\frac{1}{4}F^{a\,\mu\nu}F^a_{\mu\nu} + i\sum_{i=1}^{N_+}\overline{\psi}_+^{\,i}\not{D}^{\,(+)}\psi_+^i + i\sum_{j=1}^{N_-}\overline{\psi}_-^{\,j}\not{D}^{\,(-)}\psi_-^j, \tag{351}$$

where the chiral fermions $\psi^i_\pm$ transform according to the representations $\tau^a_{i,\pm}$ of the gauge group G ($a = 1,\ldots,\dim G$). The covariant derivatives $D^{(\pm)}_\mu$ are then defined by

$$D^{(\pm)}_\mu\psi^i_\pm = \partial_\mu\psi^i_\pm + igA^K_\mu\tau^K_{i,\pm}\psi^i_\pm. \tag{352}$$

As for global symmetries, anomalies in the gauge symmetry appear in the triangle diagram with one axial and two vector gauge current vertices

$$\langle 0|T\left[j_A^{a\,\mu}(x)j_V^{b\,\nu}(x')j_V^{c\,\sigma}(0)\right]|0\rangle = \Bigg[\quad\Bigg]_{\text{symmetric}} \tag{353}$$



[17] An early computation of the triangle diagram for the electromagnetic decay of the pion was made by Steinberger in [31].

where gauge vector and axial currents $j_V^{a\,\mu}$, $j_A^{a\,\mu}$ are given by

$$\begin{aligned} j_V^{a\mu} &= \sum_{i=1}^{N_+} \overline{\psi}_+^i \tau_+^a \gamma^\mu \psi_+^i + \sum_{j=1}^{N_-} \overline{\psi}_-^j \tau_-^a \gamma^\mu \psi_-^j, \\ j_A^{a\mu} &= \sum_{i=1}^{N_+} \overline{\psi}_+^i \tau_+^a \gamma^\mu \psi_+^i - \sum_{i=1}^{N_-} \overline{\psi}_-^j \tau_-^a \gamma^\mu \psi_-^j. \end{aligned} \tag{354}$$

Luckily, we do not have to compute the whole diagram in order to find an anomaly cancellation condition, it is enough if we calculate the overall group theoretical factor. In the case of the diagram in Eq. (353) for every fermion species running in the loop this factor is equal to

$$\mathrm{tr}\,\left[\tau_{i,\pm}^a \{\tau_{i,\pm}^b, \tau_{i,\pm}^c\}\right], \tag{355}$$

where the sign $\pm$ corresponds respectively to the generators of the representation of the gauge group for the left and right-handed fermions. Hence the anomaly cancellation condition reads

$$\sum_{i=1}^{N_+} \mathrm{tr}\,\left[\tau_{i,+}^a \{\tau_{i,+}^b, \tau_{i,+}^c\}\right] - \sum_{j=1}^{N_-} \mathrm{tr}\,\left[\tau_{j,-}^a \{\tau_{j,-}^b, \tau_{j,-}^c\}\right] = 0. \tag{356}$$

Knowing this we can proceed to check the anomaly cancellation in the standard model SU(3)×SU(2)× Left handed fermions (both leptons and quarks) transform as doublets with respect to the SU(2) factor whereas the right-handed components are singlets. The charge with respect to the U(1) part, the hypercharge Y, is determined by the Gell-Mann-Nishijima formula

$$Q = T_3 + Y, \tag{357}$$

where Q is the electric charge of the corresponding particle and T_3 is the eigenvalue with respect to the third generator of the SU(2) group in the corresponding representation: $T_3 = \frac{1}{2}\sigma^3$ for the doublets and $T_3 = 0$ for the singlets. For the first family of quarks (u, d) and leptons (e, ν_e) we have the following field content

$$\begin{array}{lccc} \text{quarks:} & \begin{pmatrix} u^\alpha \\ d^\alpha \end{pmatrix}_{L,\frac{1}{6}} & u^\alpha_{R,\frac{2}{3}} & d^\alpha_{R,\frac{2}{3}} \\ \text{leptons:} & \begin{pmatrix} \nu_e \\ e \end{pmatrix}_{L,-\frac{1}{2}} & e_{R,-1} & \end{array} \tag{358}$$

where $\alpha = 1, 2, 3$ labels the color quantum number and the subscript indicates the value of the weak hypercharge Y. Denoting the representations of SU(3)×SU(2)×U(1) by $(n_c, n_w)_Y$, with n_c and n_w the representations of SU(3) and SU(2) respectively and Y the hypercharge, the matter content of the standard model consists of a three family replication of the representations:

$$\begin{array}{lccc} \text{left-handed fermions:} & (3,2)^L_{\frac{1}{6}} & (1,2)^L_{-\frac{1}{2}} & \\ \text{right-handed fermions:} & (3,1)^R_{\frac{2}{3}} & (3,1)^R_{-\frac{1}{3}} & (1,1)^R_{-1}. \end{array} \tag{359}$$

In computing the triangle diagram we have 10 possibilities depending on which factor of the gauge group

SU(3)×SU(2)×U(1) couples to each vertex:

$SU(3)^3$	$SU(2)^3$	$U(1)^3$
$SU(3)^2\,SU(2)$	$SU(2)^2\,U(1)$	
$SU(3)^2\,U(1)$	$SU(2)\,U(1)^2$	
$SU(3)\,SU(2)^2$		
$SU(3)\,SU(2)\,U(1)$		
$SU(3)\,U(1)^2$		

It is easy to check that some of them do not give rise to anomalies. For example the anomaly for the $SU(3)^3$ case cancels because left and right-handed quarks transform in the same representation. In the case of $SU(2)^3$ the cancellation happens term by term because of the Pauli matrices identity $\sigma^a\sigma^b = \delta^{ab} + i\varepsilon^{abc}\sigma^c$ that leads to

$$\mathrm{tr}\,\left[\sigma^a\{\sigma^b,\sigma^c\}\right] = 2\,(\mathrm{tr}\,\sigma^a)\,\delta^{bc} = 0. \tag{360}$$

However the hardest anomaly cancellation condition to satisfy is the one with three U(1)'s. In this case the absence of anomalies within a single family is guaranteed by the nontrivial identity

$$\begin{aligned}\sum_{\text{left}} Y_+^3 - \sum_{\text{right}} Y_-^3 &= 3\times 2\times\left(\frac{1}{6}\right)^3 + 2\times\left(-\frac{1}{2}\right)^3 - 3\times\left(\frac{2}{3}\right)^3 - 3\times\left(-\frac{1}{3}\right)^3 - (-1)^3 \\ &= \left(-\frac{3}{4}\right) + \left(\frac{3}{4}\right) = 0.\end{aligned} \tag{361}$$

It is remarkable that the anomaly exactly cancels between leptons and quarks. Notice that this result holds even if a right-handed sterile neutrino is added since such a particle is a singlet under the whole standard model gauge group and therefore does not contribute to the triangle diagram. Therefore we see how the matter content of the standard model conspires to yield a consistent quantum field theory.

In all our discussion of anomalies we only considered the computation of one-loop diagrams. It may happen that higher loop orders impose additional conditions. Fortunately this is not so: the Adler-Bardeen theorem [35] guarantees that the axial anomaly only receives contributions from one loop diagrams. Therefore, once anomalies are canceled (if possible) at one loop we know that there will be no new conditions coming from higher-loop diagrams in perturbation theory.

The Adler-Bardeen theorem, however, only applies in perturbation theory. It is nonetheless possible that nonperturbative effects can result in the quantum violation of a gauge symmetry. This is precisely the case pointed out by Witten [36] with respect to the SU(2) gauge symmetry of the standard model. In this case the problem lies in the nontrivial topology of the gauge group SU(2). The invariance of the theory with respect to gauge transformations which are not in the connected component of the identity makes all correlation functions equal to zero. Only when the number of left-handed SU(2) fermion doublets is even gauge invariance allows for a nontrivial theory. It is again remarkable that the family structure of the standard model makes this anomaly to cancel

$$3\times\begin{pmatrix} u \\ d \end{pmatrix}_L + 1\times\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L = 4\ \text{SU(2)-doublets}, \tag{362}$$

where the factor of 3 comes from the number of colors.

8 Renormalization

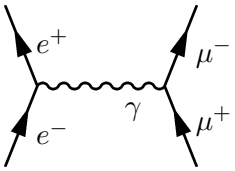
8.1 Removing infinities

From its very early stages, quantum field theory was faced with infinities. They emerged in the calculation of most physical quantities, such as the correction to the charge of the electron due to the interactions with the radiation field. The way these divergences where handled in the 1940s, starting with Kramers, was physically very much in the spirit of the Quantum Theory emphasis in observable quantities: since the observed magnitude of physical quantities (such as the charge of the electron) is finite, this number should arise from the addition of a "bare" (unobservable) value and the quantum corrections. The fact that both of these quantities were divergent was not a problem physically, since only its finite sum was an observable quantity. To make thing mathematically sound, the handling of infinities requires the introduction of some regularization procedure which cuts the divergent integrals off at some momentum scale Λ. Morally speaking, the physical value of an observable $\mathcal{O}_{\rm physical}$ is given by

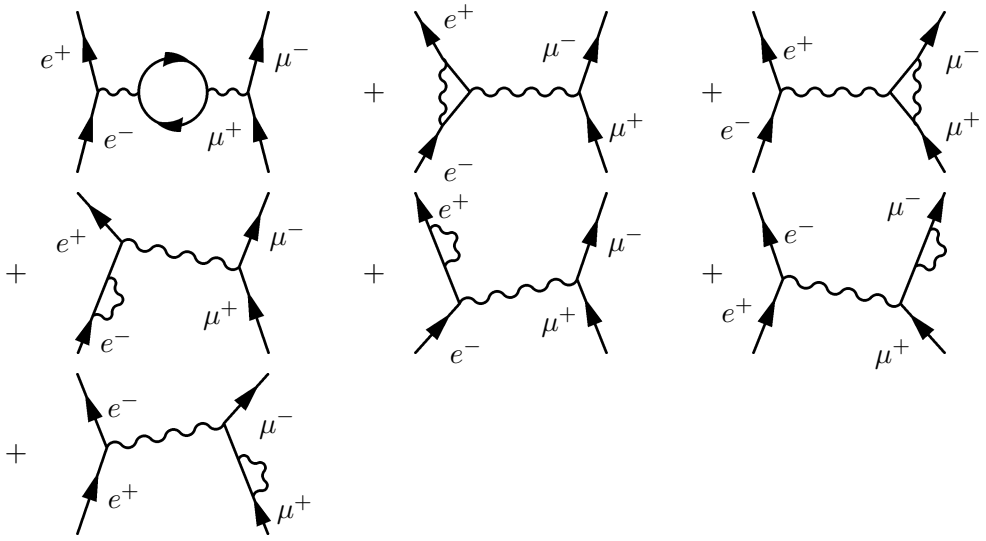
$$\mathcal{O}_{\rm physical} = \lim_{\Lambda\to\infty} \left[\mathcal{O}(\Lambda)_{\rm bare} + \Delta\mathcal{O}(\Lambda)_\hbar\right], \tag{363}$$

where $\Delta\mathcal{O}(\Lambda)_\hbar$ represents the regularized quantum corrections.

To make this qualitative discussion more precise we compute the corrections to the electric charge in Quantum Electrodynamics. We consider the process of annihilation of an electron-positron pair to create a muon-antimuon pair $e^-e^+ \to \mu^+\mu^-$. To lowest order in the electric charge e the only diagram contributing is

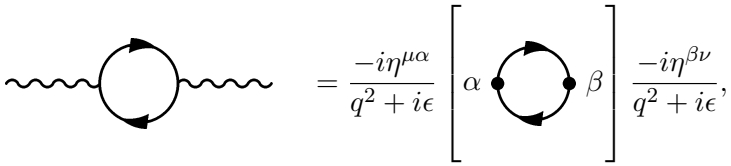


However, the corrections at order e^4 to this result requires the calculation of seven more diagrams



In order to compute the renormalization of the charge we consider the first diagram which takes into account the first correction to the propagator of the virtual photon interchanged between the pairs

due to vacuum polarization. We begin by evaluating



$$= \frac{-i\eta^{\mu\alpha}}{q^2+i\epsilon}\left[\alpha \quad \beta\right]\frac{-i\eta^{\beta\nu}}{q^2+i\epsilon}, \tag{364}$$

where the diagram between brackets is given by

$$\equiv \Pi^{\alpha\beta}(q) = i^2(-ie)^2(-1)\int \frac{d^4k}{(2\pi)^4}\frac{\mathrm{Tr}\,(\not{k}+m_e)\gamma^\alpha(\not{k}+\not{q}+m_e)\gamma^\beta}{[k^2-m_e^2+i\epsilon]\,[(k+q)^2-m_e^2+i\epsilon]}. \tag{365}$$

Physically this diagram includes the correction to the propagator due to the polarization of the vacuum, i.e. the creation of virtual electron-positron pairs by the propagating photon. The momentum q is the total momentum of the electron-positron pair in the intermediate channel.

It is instructive to look at this diagram from the point of view of perturbation theory in nonrelativistic Quantum Mechanics. In each vertex the interaction consists of the annihilation (resp. creation) of a photon and the creation (resp. annihilation) of an electron-positron pair. This can be implemented by the interaction Hamiltonian

$$H_{\rm int} = e\int d^3x\,\overline{\psi}\gamma^\mu\psi A_\mu. \tag{366}$$

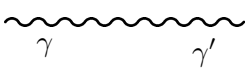
All fields inside the integral can be expressed in terms of the corresponding creation-annihilation operators for photons, electrons and positrons. In Quantum Mechanics, the change in the wave function at first order in the perturbation $H_{\rm int}$ is given by

$$|\gamma,{\rm in}\rangle = |\gamma,{\rm in}\rangle_0 + \sum_n \frac{\langle n|H_{\rm int}|\gamma,{\rm in}\rangle_0}{E_{\rm in}-E_n}|n\rangle \tag{367}$$

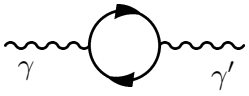
and similarly for $|\gamma,{\rm out}\rangle$, where we have denoted symbolically by $|n\rangle$ all the possible states of the electron-positron pair. Since these states are orthogonal to $|\gamma,{\rm in}\rangle_0$, $|\gamma,{\rm out}\rangle_0$, we find torder e^2

$$\langle\gamma,{\rm in}|\gamma',{\rm out}\rangle = {}_0\langle\gamma,{\rm in}|\gamma',{\rm out}\rangle_0 + \sum_n \frac{{}_0\langle\gamma,{\rm in}|H_{\rm int}|n\rangle\,\langle n|H_{\rm int}|\gamma',{\rm out}\rangle_0}{(E_{\rm in}-E_n)(E_{\rm out}-E_n)} + \mathcal{O}(e^4). \tag{368}$$

Hence, we see that the diagram of Eq. (364) really corresponds to the order-e^2 correction to the photon propagator $\langle\gamma,{\rm in}|\gamma',{\rm out}\rangle$



$$\longrightarrow \quad {}_0\langle\gamma,{\rm in}|\gamma',{\rm out}\rangle_0$$



$$\longrightarrow \quad \sum_n \frac{\langle\gamma,{\rm in}|H_{\rm int}|n\rangle\,\langle n|H_{\rm int}|\gamma',{\rm out}\rangle}{(E_{\rm in}-E_n)(E_{\rm out}-E_n)}. \tag{369}$$

Once we understood the physical meaning of the Feynman diagram to be computed we proceed to its evaluation. In principle there is no problem in computing the integral in Eq. (364) for nonzero values of the electron mass. However since here we are going to be mostly interested in seeing how the divergence of the integral results in a scale-dependent renormalization of the electric charge, we will set $m_e = 0$. This is something safe to do, since in the case of this diagram we are not inducing new infrared divergences in taking the electron as massless. Implementing gauge invariance and using standard techniques in the computation of Feynman diagrams (see references [1]- [11]) the polarization tensor $\Pi_{\mu\nu}(q)$ defined in Eq. (365) can be written as

$$\Pi_{\mu\nu}(q) = \left(q^2\eta_{\mu\nu} - q_\mu q_\nu\right)\Pi(q^2) \tag{370}$$

with

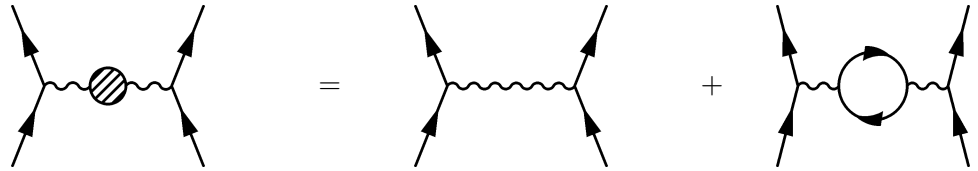
$$\Pi(q) = 8e^2\int_0^1 dx \int \frac{d^4k}{(2\pi)^4}\frac{x(1-x)}{[k^2 - m^2 + x(1-x)q^2 + i\epsilon]^2} \tag{371}$$

To handle this divergent integral we have to figure out some procedure to render it finite. This can be done in several ways, but here we choose to cut the integrals off at a high energy scale Λ, where new physics might be at work, $|p| < \Lambda$. This gives the result

$$\Pi(q^2) \simeq \frac{e^2}{12\pi^2}\log\left(\frac{q^2}{\Lambda^2}\right) + \text{finite terms.} \tag{372}$$

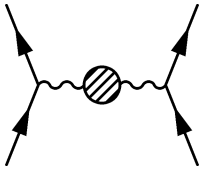
If we would send the cutoff to infinity $\Lambda \to \infty$ the divergence blows up and something has to be done about it.

If we want to make sense out of this, we have to go back to the physical question that led us to compute Eq. (364). Our primordial motivation was to compute the corrections to the annihilation of two electrons into two muons. Including the correction to the propagator of the virtual photon we have



$$\begin{aligned} &= \eta_{\alpha\beta}\left(\overline{v}_e\gamma^\alpha u_e\right)\frac{e^2}{4\pi q^2}\left(\overline{v}_\mu\gamma^\beta u_\mu\right) + \eta_{\alpha\beta}\left(\overline{v}_e\gamma^\alpha u_e\right)\frac{e^2}{4\pi q^2}\Pi(q^2)\left(\overline{v}_\mu\gamma^\beta u_\mu\right) \\ &= \eta_{\alpha\beta}\left(\overline{v}_e\gamma^\alpha u_e\right)\left\{\frac{e^2}{4\pi q^2}\left[1 + \frac{e^2}{12\pi^2}\log\left(\frac{q^2}{\Lambda^2}\right)\right]\right\}\left(\overline{v}_\mu\gamma^\beta u_\mu\right). \end{aligned} \tag{373}$$

Now let us imagine that we are performing a $e^-\,e^+ \to \mu^-\mu^+$ with a center of mass energy μ. From the previous result we can identify the effective charge of the particles at this energy scale $e(\mu)$ as



$$= \eta_{\alpha\beta}\left(\overline{v}_e\gamma^\alpha u_e\right)\left[\frac{e(\mu)^2}{4\pi q^2}\right]\left(\overline{v}_\mu\gamma^\beta u_\mu\right). \tag{374}$$

This charge, $e(\mu)$, is the quantity that is physically measurable in our experiment. Now we can make sense of the formally divergent result (373) by assuming that the charge appearing in the classical Lagrangian of QED is just a "bare" value that depends on the scale Λ at which we cut off the theory, $e \equiv e(\Lambda)_{\text{bare}}$. In order to reconcile (373) with the physical results (374) we must assume that the dependence of the bare (unobservable) charge $e(\Lambda)_{\text{bare}}$ on the cutoff Λ is determined by the identity

$$e(\mu)^2 = e(\Lambda)^2_{\text{bare}}\left[1 + \frac{e(\Lambda)^2_{\text{bare}}}{12\pi^2}\log\left(\frac{\mu^2}{\Lambda^2}\right)\right]. \tag{375}$$

If we still insist in removing the cutoff, $\Lambda \to \infty$ we have to send the bare charge to zero $e(\Lambda)_{\rm bare} \to 0$ in such a way that the effective coupling has the finite value given by the experiment at the energy scale μ. It is not a problem, however, that the bare charge is small for large values of the cutoff, since the only measurable quantity is the effective charge that remains finite. Therefore all observable quantities should be expressed in perturbation theory as a power series in the physical coupling $e(\mu)^2$ and not in the unphysical bare coupling $e(\Lambda)_{\rm bare}$.

8.2 The beta-function and asymptotic freedom

We can look at the previous discussion, an in particular Eq. (375), from a different point of view. In order to remove the ambiguities associated with infinities we have been forced to introduce a dependence of the coupling constant on the energy scale at which a process takes place. From the expression of the physical coupling in terms of the bare charge (375) we can actually eliminate the cutoff Λ, whose value after all should not affect the value of physical quantities. Taking into account that we are working in perturbation theory in $e(\mu)^2$, we can express the bare charge $e(\Lambda)^2_{\rm bare}$ in terms of $e(\mu)^2$ as

$$e(\Lambda)^2 = e(\mu)^2 \left[1 + \frac{e(\mu)^2}{12\pi^2} \log\left(\frac{\mu^2}{\Lambda^2}\right)\right] + \mathcal{O}[e(\mu)^6]. \tag{376}$$

This expression allow us to eliminate all dependence in the cutoff in the expression of the effective charge at a scale μ by replacing $e(\Lambda)_{\rm bare}$ in Eq. (375) by the one computed using (376) at a given reference energy scale μ_0

$$e(\mu)^2 = e(\mu_0)^2 \left[1 + \frac{e(\mu_0)^2}{12\pi^2} \log\left(\frac{\mu^2}{\mu_0^2}\right)\right]. \tag{377}$$

From this equation we can compute, at this order in perturbation theory, the effective value of the coupling constant at an energy μ, once we know its value at some reference energy scale μ_0. In the case of the electron charge we can use as a reference Thompson's scattering at energies of the order of the electron mass $m_e \simeq 0.5$ MeV, at where the value of the electron charge is given by the well known value

$$e(m_e)^2 \simeq \frac{1}{137}. \tag{378}$$

With this we can compute $e(\mu)^2$ at any other energy scale applying Eq. (377), for example at the electron mass $\mu = m_e \simeq 0.5\,$MeV. However, in computing the electromagnetic coupling constant at any other scale we must take into account the fact that other charged particles can run in the loop in Eq. (373). Suppose, for example, that we want to calculate the fine structure constant at the mass of the Z^0-boson $\mu = M_Z \equiv 92$ GeV. Then we should include in Eq. (377) the effect of other fermionic standard model fields with masses below M_Z. Doing this, we find[18]

$$e(M_Z)^2 = e(m_e)^2 \left[1 + \frac{e(m_e)^2}{12\pi^2} \left(\sum_i q_i^2\right) \log\left(\frac{M_Z^2}{m_e^2}\right)\right], \tag{379}$$

where q_i is the charge in units of the electron charge of the i-th fermionic species running in the loop and we sum over all fermions with masses below the mass of the Z^0 boson. This expression shows how the electromagnetic coupling grows with energy. However, in order to compare with the experimental value of $e(M_Z)^2$ it is not enough with including the effect of fermionic fields, since also the $W^\pm$ bosons

[18] In the first version of these notes the argument used to show the growing of the electromagnetic coupling constant could have led to confusion to some readers. To avoid this potential problem we include in the equation for the running coupling $e(\mu)^2$ the contribution of all fermions with masses below M_Z. We thank Lubos Motl for bringing this issue to our attention.

can run in the loop ($M_W < M_Z$). Taking this into account, as well as threshold effects, the value of the electron charge at the scale M_Z is found to be [37]

$$e(M_Z)^2 \simeq \frac{1}{128.9} \ . \tag{380}$$

This growing of the effective fine structure constant with energy can be understood heuristically by remembering that the effect of the polarization of the vacuum shown in the diagram of Eq. (364) amounts to the creation of a plethora of electron-positron pairs around the location of the charge. These virtual pairs behave as dipoles that, as in a dielectric medium, tend to screen this charge and decreasing its value at long distances (i.e. lower energies).

The variation of the coupling constant with energy is usually encoded in quantum field theory in the *beta function* defined by

$$\beta(g) = \mu \frac{dg}{d\mu}. \tag{381}$$

In the case of QED the beta function can be computed from Eq. (377) with the result

$$\beta(e)_{\rm QED} = \frac{e^3}{12\pi^2}. \tag{382}$$

The fact that the coefficient of the leading term in the beta-function is positive $\beta_0 \equiv \frac{1}{6\pi} > 0$ gives us the overall behavior of the coupling as we change the scale. Eq. (382) means that, if we start at an energy where the electric coupling is small enough for our perturbative treatment to be valid, the effective charge grows with the energy scale. This growing of the effective coupling constant with energy means that QED is infrared safe, since the perturbative approximation gives better and better results as we go to lower energies. Actually, because the electron is the lighter electrically charged particle and has a finite nonvanishing mass the running of the fine structure constant stops at the scale m_e in the well-known value $\frac{1}{137}$. Would other charged fermions with masses below m_e be present in Nature, the effective value of the fine structure constant in the interaction between these particles would run further to lower values at energies below the electron mass.

On the other hand if we increase the energy scale $e(\mu)^2$ grows until at some scale the coupling is of order one and the perturbative approximation breaks down. In QED this is known as the problem of the Landau pole but in fact it does not pose any serious threat to the reliability of QED perturbation theory: a simple calculation shows that the energy scale at which the theory would become strongly coupled is $\Lambda_{\rm Landau} \simeq 10^{277}$ GeV. However, we know that QED does not live that long! At much lower scales we expect electromagnetism to be unified with other interactions, and even if this is not the case we will enter the uncharted territory of quantum gravity at energies of the order of 10^{19} GeV.

So much for QED. The next question that one may ask at this stage is whether it is possible to find quantum field theories with a behavior opposite to that of QED, i.e. such that they become weakly coupled at high energies. This is not a purely academic question. In the late 1960s a series of deep-inelastic scattering experiments carried out at SLAC showed that the quarks behave essentially as free particles inside hadrons. The apparent problem was that no theory was known at that time that would become free at very short distances: the example set by QED seem to be followed by all the theories that were studied. This posed a very serious problem for quantum field theory as a way to describe subnuclear physics, since it seemed that its predictive power was restricted to electrodynamics but failed miserably when applied to describe strong interactions.

Nevertheless, this critical time for quantum field theory turned out to be its finest hour. In 1973 David Gross and Frank Wilczek [38] and David Politzer [39] showed that nonabelian gauge theories can actually display the required behavior. For the QCD Lagrangian in Eq. (331) the beta function is given

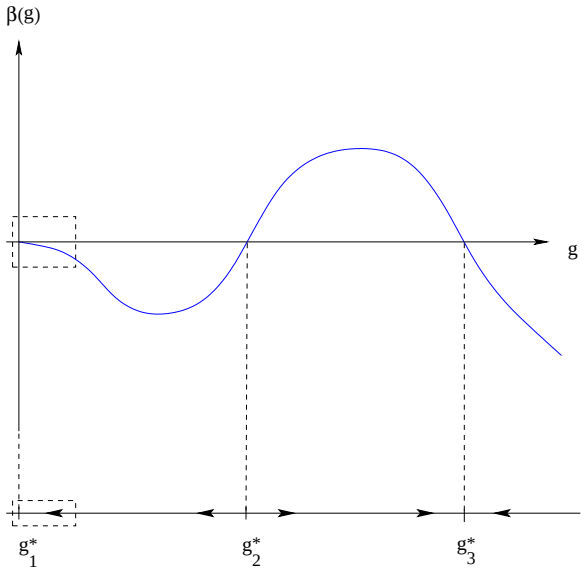


Fig. 14: Beta function for a hypothetical theory with three fixed points g_1^*, g_2^* and g_3^*. A perturbative analysis would capture only the regions shown in the boxes.

by[19]

$$\beta(g) = -\frac{g^3}{16\pi^2}\left[\frac{11}{3}N_c - \frac{2}{3}N_f\right]. \tag{383}$$

In particular, for real QCD ($N_C = 3$, $N_f = 6$) we have that $\beta(g) = -\frac{7g^3}{16\pi^2} < 0$. This means that for a theory that is weakly coupled at an energy scale μ_0 the coupling constant decreases as the energy increases $\mu \to \infty$. This explain the apparent freedom of quarks inside the hadrons: when the quarks are very close together their effective color charge tend to zero. This phenomenon is called *asymptotic freedom*.

Asymptotic free theories display a behavior that is opposite to that found above in QED. At high energies their coupling constant approaches zero whereas at low energies they become strongly coupled (infrared slavery). This features are at the heart of the success of QCD as a theory of strong interactions, since this is exactly the type of behavior found in quarks: they are quasi-free particles inside the hadrons but the interaction potential potential between them increases at large distances.

Although asymptotic free theories can be handled in the ultraviolet, they become extremely complicated in the infrared. In the case of QCD it is still to be understood (at least analytically) how the theory confines color charges and generates the spectrum of hadrons, as well as the breaking of the chiral symmetry (345).

In general, the ultraviolet and infrared properties of a theory are controlled by the fixed points of the beta function, i.e. those values of the coupling constant g for which it vanishes

$$\beta(g^*) = 0. \tag{384}$$

Using perturbation theory we have seen that for both QED and QCD one of such fixed points occurs at zero coupling, $g^* = 0$. However, our analysis also showed that the two theories present radically different behavior at high and low energies. From the point of view of the beta function, the difference lies in the energy regime at which the coupling constant approaches its critical value. This is in fact governed by the sign of the beta function around the critical coupling.

[19]The expression of the beta function of QCD was also known to 't Hooft [40]. There are even earlier computations in the russian literature [41].

We have seen above that when the beta function is negative close to the fixed point (the case of QCD) the coupling tends to its critical value, $g^* = 0$, as the energy is increased. This means that the critical point is *ultraviolet stable*, i.e. it is an attractor as we evolve towards higher energies. If, on the contrary, the beta function is positive (as it happens in QED) the coupling constant approaches the critical value as the energy decreases. This is the case of an *infrared stable* fixed point.

This analysis that we have motivated with the examples of QED and QCD is completely general and can be carried out for any quantum field theory. In Fig. 14 we have represented the beta function for a hypothetical theory with three fixed points located at couplings g_1^*, g_2^* and g_3^*. The arrows in the line below the plot represent the evolution of the coupling constant as the energy increases. From the analysis presented above we see that $g_1^* = 0$ and g_3^* are ultraviolet stable fixed points, while the fixed point g_2^* is infrared stable.

In order to understand the high and low energy behavior of a quantum field theory it is then crucial to know the structure of the beta functions associated with its couplings. This can be a very difficult task, since perturbation theory only allows the study of the theory around “trivial" fixed points, i.e. those that occur at zero coupling like the case of g_1^* in Fig. 14. On the other hand, any “nontrivial” fixed point occurring in a theory (like g_2^* and g_3^*) cannot be captured in perturbation theory and requires a full nonperturbative analysis.

The moral to be learned from our discussion above is that dealing with the ultraviolet divergences in a quantum field theory has the consequence, among others, of introducing an energy dependence in the measured value of the coupling constants of the theory (for example the electric charge in QED). This happens even in the case of renormalizable theories without mass terms. These theories are scale invariant at the classical level because the action does not contain any dimensionful parameter. In this case the running of the coupling constants can be seen as resulting from a quantum breaking of classical scale invariance: different energy scales in the theory are distinguished by different values of the coupling constants. Remembering what we learned in Section 7, we conclude that classical scale invariance is an anomalous symmetry. One heuristic way to see how the conformal anomaly comes about is to notice that the regularization of an otherwise scale invariant field theory requires the introduction of an energy scale (e.g. a cutoff). This breaking of scale invariance cannot be restored after renormalization.

Nevertheless, scale invariance is not lost forever in the quantum theory. It is recovered at the fixed points of the beta function where, by definition, the coupling does not run. To understand how this happens we go back to a scale invariant classical field theory whose field $\phi(x)$ transform under coordinate rescalings as

$$x^\mu \longrightarrow \lambda x^\mu, \qquad \phi(x) \longrightarrow \lambda^{-\Delta}\phi(\lambda^{-1}x), \tag{385}$$

where Δ is called the canonical scaling dimension of the field. An example of such a theory is a massless ϕ^4 theory in four dimensions

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{g}{4!}\phi^4, \tag{386}$$

where the scalar field has canonical scaling dimension $\Delta = 1$. The Lagrangian density transforms as

$$\mathcal{L} \longrightarrow \lambda^{-4}\mathcal{L}[\phi] \tag{387}$$

and the classical action remains invariant[20].

If scale invariance is preserved under quantization, the Green’s functions transform as

$$\langle\Omega|T[\phi'(x_1)\dots\phi'(x_n)]|\Omega\rangle = \lambda^{n\Lambda}\langle\Omega|T[\phi(\lambda^{-1}x_1)\dots\phi(\lambda^{-1}x_n)]|\Omega\rangle. \tag{388}$$

[20]In a D-dimensional theory the canonical scaling dimensions of the fields coincide with its engineering dimension: $\Delta = \frac{D-2}{2}$ for bosonic fields and $\Delta = \frac{D-1}{2}$ for fermionic ones. For a Lagrangian with no dimensionful parameters classical scale invariance follows then from dimensional analysis.

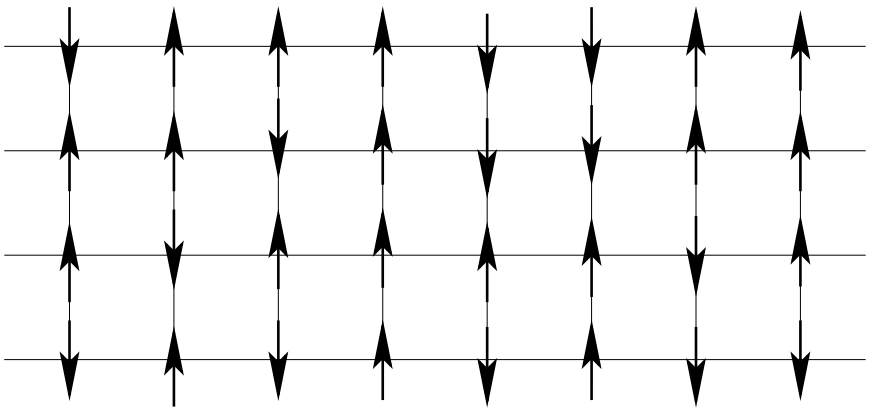


Fig. 15: Systems of spins in a two-dimensional square lattice.

This is precisely what happens in a free theory. In an interacting theory the running of the coupling constant destroys classical scale invariance at the quantum level. Despite of this, at the fixed points of the beta function the Green's functions transform again according to (388) where Δ is replaced by

$$\Delta_{\rm anom} = \Delta + \gamma^*. \tag{389}$$

The canonical scaling dimension of the fields are corrected by γ^*, which is called the anomalous dimension. They carry the dynamical information about the high-energy behavior of the theory.

8.3 The renormalization group

In spite of its successes, the renormalization procedure presented above can be seen as some kind of prescription or recipe to get rid of the divergences in an ordered way. This discomfort about renormalization was expressed in occasions by comparing it with "sweeping the infinities under the rug". However thanks to Ken Wilson to a large extent [42] the process of renormalization is now understood in a very profound way as a procedure to incorporate the effects of physics at high energies by modifying the value of the parameters that appear in the Lagrangian.

Statistical mechanics. Wilson's ideas are both simple and profound and consist in thinking about quantum field theory as the analog of a thermodynamical description of a statistical system. To be more precise, let us consider an Ising spin system in a two-dimensional square lattice as the one depicted in Fig 15. In terms of the spin variables $s_i = \pm\frac{1}{2}$, where i labels the lattice site, the Hamiltonian of the system is given by

$$H = -J \sum_{\langle i,j \rangle} s_i \, s_j, \tag{390}$$

where $\langle i, j\rangle$ indicates that the sum extends over nearest neighbors and J is the coupling constant between neighboring spins (here we consider that there is no external magnetic field). The starting point to study the statistical mechanics of this system is the partition function defined as

$$\mathcal{Z} = \sum_{\{s_i\}} e^{-\beta H}, \tag{391}$$

where the sum is over all possible configurations of the spins and $\beta = \frac{1}{T}$ is the inverse temperature. For $J > 0$ the Ising model presents spontaneous magnetization below a critical temperature T_c, in any dimension higher than one. Away from this temperature correlations between spins decay exponentially at large distances

$$\langle s_i s_j \rangle \sim e^{-\frac{|x_{ij}|}{\xi}}, \tag{392}$$

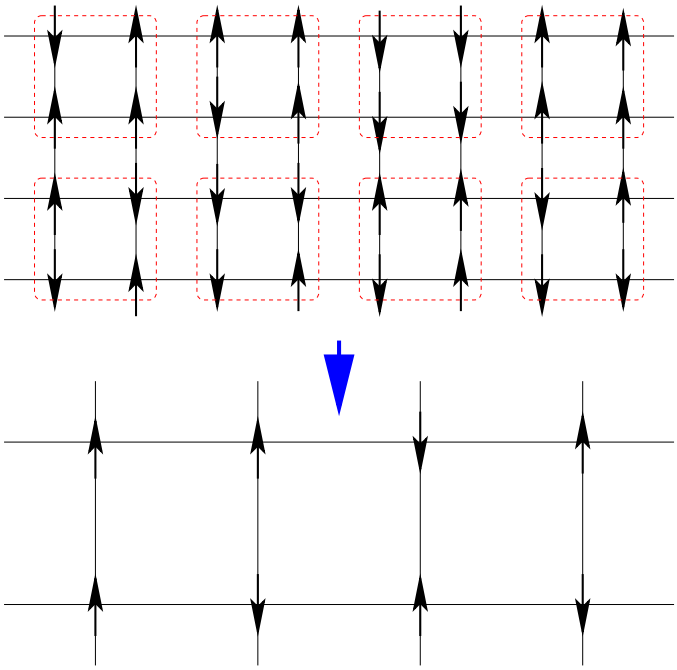


Fig. 16: Decimation of the spin lattice. Each block in the upper lattice is replaced by an effective spin computed according to the rule (394). Notice also that the size of the lattice spacing is doubled in the process.

with $|x_{ij}|$ the distance between the spins located in the i-th and j-th sites of the lattice. This expression serves as a definition of the correlation length ξ which sets the characteristic length scale at which spins can influence each other by their interaction through their nearest neighbors.

Suppose now that we are interested in a macroscopic description of this spin system. We can capture the relevant physics by integrating out somehow the physics at short scales. A way in which this can be done was proposed by Leo Kadanoff [43] and consists in dividing our spin system in spin-blocks like the ones showed in Fig 16. Now we can construct another spin system where each spin-block of the original lattice is replaced by an effective spin calculated according to some rule from the spins contained in each block B_a

$$\{s_i : i \in B_a\} \quad \longrightarrow \quad {s_a}^{(1)}. \tag{393}$$

For example we can define the effective spin associated with the block B_a by taking the majority rule with an additional prescription in case of a draw

$${s_a}^{(1)} = \frac{1}{2}\text{sgn}\left(\sum_{i\in B_a} s_i\right), \tag{394}$$

where we have used the sign function, $\text{sign}(x) \equiv \frac{x}{|x|}$, with the additional definition $\text{sgn}(0) = 1$. This procedure is called decimation and leads to a new spin system with a doubled lattice space.

The idea now is to rewrite the partition function (391) only in terms of the new effective spins ${s_a}^{(1)}$. Then we start by splitting the sum over spin configurations into two nested sums, one over the spin blocks and a second one over the spins within each block

$$\mathcal{Z} = \sum_{\{\vec{s}\}} e^{-\beta H[s_i]} = \sum_{\{\vec{s}^{(1)}\}} \sum_{\{\vec{s}\in B_a\}} \delta\left[{s_a}^{(1)} - \text{sign}\left(\sum_{i\in B_a} s_i\right)\right] e^{-\beta H[s_i]}. \tag{395}$$

The interesting point now is that the sum over spins inside each block can be written as the exponential of a new effective Hamiltonian depending only on the effective spins, $H^{(1)}[s_a{}^{(1)}]$

$$\sum_{\{s\in B_a\}} \delta\left[s_a{}^{(1)} - \text{sign}\left(\sum_{i\in B_a} s_i\right)\right] e^{-\beta H[s_i]} = e^{-\beta H^{(1)}[s_a^{(1)}]}. \tag{396}$$

The new Hamiltonian is of course more complicated

$$H^{(1)} = -J^{(1)} \sum_{\langle i,j\rangle} s_i^{(1)} s_j^{(1)} + \dots \tag{397}$$

where the dots stand for other interaction terms between the effective block spins. This new terms appear because in the process of integrating out short distance physics we induce interactions between the new effective degrees of freedom. For example the interaction between the spin block variables $s_i^{(1)}$ will in general not be restricted to nearest neighbors in the new lattice. The important point is that we have managed to rewrite the partition function solely in terms of this new (renormalized) spin variables $s^{(1)}$ interacting through a new Hamiltonian $H^{(1)}$

$$\mathcal{Z} = \sum_{\{s^{(1)}\}} e^{-\beta H^{(1)}[s_a^{(1)}]}. \tag{398}$$

Let us now think about the space of all possible Hamiltonians for our statistical system including all kinds of possible couplings between the individual spins compatible with the symmetries of the system. If denote by $\mathcal{R}$ the decimation operation, our previous analysis shows that $\mathcal{R}$ defines a map in this space of Hamiltonians

$$\mathcal{R} : H \to H^{(1)}. \tag{399}$$

At the same time the operation $\mathcal{R}$ replaces a lattice with spacing a by another one with double spacing $2a$. As a consequence the correlation length in the new lattice measured in units of the lattice spacing is divided by two, $\mathcal{R} : \xi \to \frac{\xi}{2}$.

Now we can iterate the operation $\mathcal{R}$ an indefinite number of times. Eventually we might reach a Hamiltonian $H_\star$ that is not further modified by the operation $\mathcal{R}$

$$H \xrightarrow{\mathcal{R}} H^{(1)} \xrightarrow{\mathcal{R}} H^{(2)} \xrightarrow{\mathcal{R}} \dots \xrightarrow{\mathcal{R}} H_\star. \tag{400}$$

The fixed point Hamiltonian $H_\star$ is *scale invariant* because it does not change as $\mathcal{R}$ is performed. Notice that because of this invariance the correlation length of the system at the fixed point do not change under $\mathcal{R}$. This fact is compatible with the transformation $\xi \to \frac{\xi}{2}$ only if $\xi = 0$ or $\xi = \infty$. Here we will focus in the case of nontrivial fixed points with infinite correlation length.

The space of Hamiltonians can be parametrized by specifying the values of the coupling constants associated with all possible interaction terms between individual spins of the lattice. If we denote by $\mathcal{O}_a[s_i]$ these (possibly infinite) interaction terms, the most general Hamiltonian for the spin system under study can be written as

$$H[s_i] = \sum_{a=1}^{\infty} \lambda_a \mathcal{O}_a[s_i], \tag{401}$$

where $\lambda_a \in \mathbb{R}$ are the coupling constants for the corresponding operators. These constants can be thought of as coordinates in the space of all Hamiltonians. Therefore the operation $\mathcal{R}$ defines a transformation in the set of coupling constants

$$\mathcal{R} : \lambda_a \longrightarrow \lambda_a^{(1)}. \tag{402}$$

For example, in our case we started with a Hamiltonian in which only one of the coupling constants is different from zero (say $\lambda_1 = -J$). As a result of the decimation $\lambda_1 \equiv -J \to -J^{(1)}$ while some of the originally vanishing coupling constants will take a nonzero value. Of course, for the fixed point Hamiltonian the coupling constants do not change under the scale transformation $\mathcal{R}$.

Physically the transformation $\mathcal{R}$ integrates out short distance physics. The consequence for physics at long distances is that we have to replace our Hamiltonian by a new one with different values for the coupling constants. That is, our ignorance of the details of the physics going on at short distances result in a *renormalization* of the coupling constants of the Hamiltonian that describes the long range physical processes. It is important to stress that although $\mathcal{R}$ is sometimes called a renormalization group transformation in fact this is a misnomer. Transformations between Hamiltonians defined by $\mathcal{R}$ do not form a group: since these transformations proceed by integrating out degrees of freedom at short scales they cannot be inverted.

In statistical mechanics fixed points under renormalization group transformations with $\xi = \infty$ are associated with phase transitions. From our previous discussion we can conclude that the space of Hamiltonians is divided in regions corresponding to the basins of attraction of the different fixed points. We can ask ourselves now about the stability of those fixed points. Suppose we have a statistical system described by a fixed-point Hamiltonian $H_\star$ and we perturb it by changing the coupling constant associated with an interaction term $\mathcal{O}$. This is equivalent to replace $H_\star$ by the perturbed Hamiltonian

$$H = H_\star + \delta\lambda\, \mathcal{O}, \tag{403}$$

where $\delta\lambda$ is the perturbation of the coupling constant corresponding to $\mathcal{O}$ (we can also consider perturbations in more than one coupling constant). At the same time thinking of the λ_a's as coordinates in the space of all Hamiltonians this corresponds to moving slightly away from the position of the fixed point.

The question to decide now is in which direction the renormalization group flow will take the perturbed system. Working at first order in $\delta\lambda$ there are three possibilities:

- The renormalization group flow takes the system back to the fixed point. In this case the corresponding interaction $\mathcal{O}$ is called *irrelevant*.
- $\mathcal{R}$ takes the system away from the fixed point. If this is what happens the interaction is called *relevant*.
- It is possible that the perturbation actually does not take the system away from the fixed point at first order in $\delta\lambda$. In this case the interaction is said to be *marginal* and it is necessary to go to higher orders in $\delta\lambda$ in order to decide whether the system moves to or away the fixed point, or whether we have a family of fixed points.

Therefore we can picture the action of the renormalization group transformation as a flow in the space of coupling constants. In Fig. 17 we have depicted an example of such a flow in the case of a system with two coupling constants λ_1 and λ_2. In this example we find two fixed points, one at the origin O and another at F for a finite value of the couplings. The arrows indicate the direction in which the renormalization group flow acts. The free theory at $\lambda_1 = \lambda_2 = 0$ is a stable fix point since any perturbation $\delta\lambda_1, \delta\lambda_2 > 0$ makes the theory flow back to the free theory at long distances. On the other hand, the fixed point F is stable with respect to certain type of perturbations (along the line with incoming arrows) whereas for any other perturbations the system flows either to the free theory at the origin or to a theory with infinite values for the couplings.

Quantum field theory. Let us see now how these ideas of the renormalization group apply to Field Theory. Let us begin with a quantum field theory defined by the Lagrangian

$$\mathcal{L}[\phi_a] = \mathcal{L}_0[\phi_a] + \sum_i g_i \mathcal{O}_i[\phi_a], \tag{404}$$

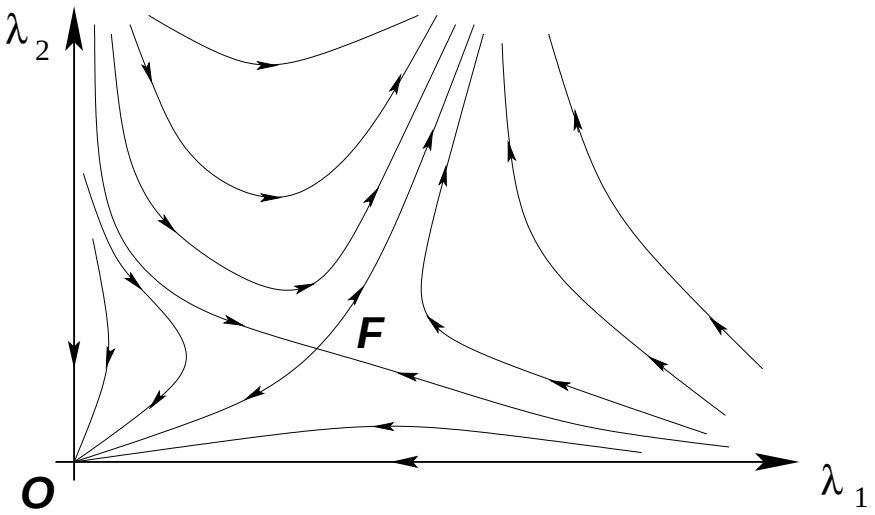


Fig. 17: Example of a renormalization group flow.

where $\mathcal{L}_0[\phi_a]$ is the kinetic part of the Lagrangian and g_i are the coupling constants associated with the operators $\mathcal{O}_i[\phi_a]$. In order to make sense of the quantum theory we introduce a cutoff in momenta Λ. In principle we include all operators $\mathcal{O}_i$ compatible with the symmetries of the theory.

In section 8.2 we saw how in the cases of QED and QCD, the value of the coupling constant changed with the scale from its value at the scale Λ. We can understand now this behavior along the lines of the analysis presented above for the Ising model. If we would like to compute the effective dynamics of the theory at an energy scale $\mu < \Lambda$ we only have to integrate out all physical models with energies between the cutoff Λ and the scale of interest μ. This is analogous to what we did in the Ising model by replacing the original spins by the block spins. In the case of field theory the effective action $S[\phi_a, \mu]$ at scale μ can be written in the language of functional integration as

$$e^{iS[\phi'_a,\mu]} = \int_{\mu<p<\Lambda} \prod_a \mathcal{D}\phi_a \, e^{iS[\phi_a,\Lambda]}. \tag{405}$$

Here $S[\phi_a, \Lambda]$ is the action at the cutoff scale

$$S[\phi_a, \Lambda] = \int d^4x \left\{ \mathcal{L}_0[\phi_a] + \sum_i g_i(\Lambda)\mathcal{O}_i[\phi_a] \right\} \tag{406}$$

and the functional integral in Eq. (405) is carried out only over the field modes with momenta in the range $\mu < p < \Lambda$. The action resulting from integrating out the physics at the intermediate scales between Λ and μ depends not on the original field variable ϕ_a but on some renormalized field ϕ'_a. At the same time the couplings $g_i(\mu)$ differ from their values at the cutoff scale $g_i(\Lambda)$. This is analogous to what we learned in the Ising model: by integrating out short distance physics we ended up with a new Hamiltonian depending on renormalized effective spin variables and with renormalized values for the coupling constants. Therefore the resulting effective action at scale μ can be written as

$$S[\phi'_a, \mu] = \int d^4x \left\{ \mathcal{L}_0[\phi'_a] + \sum_i g_i(\mu)\mathcal{O}_i[\phi'_a] \right\}. \tag{407}$$

This Wilsonian interpretation of renormalization sheds light to what in section 8.1 might have looked just a smart way to get rid of the infinities. The running of the coupling constant with the energy scale can be understood now as a way of incorporating into an effective action at scale μ the effects of field excitations at higher energies $E > \mu$.

As in statistical mechanics there are also quantum field theories that are fixed points of the renormalization group flow, i.e. whose coupling constants do not change with the scale. We have encountered them already in Section 8.2 when studying the properties of the beta function. The most trivial example of such theories are massless free quantum field theories, but there are also examples of four-dimensional interacting quantum field theories which are scale invariant. Again we can ask the question of what happens when a scale invariant theory is perturbed with some operator. In general the perturbed theory is not scale invariant anymore but we may wonder whether the perturbed theory flows at low energies towards or away the theory at the fixed point.

In quantum field theory this can be decided by looking at the canonical dimension $d[\mathcal{O}]$ of the operator $\mathcal{O}[\phi_a]$ used to perturb the theory at the fixed point. In four dimensions the three possibilities are defined by:

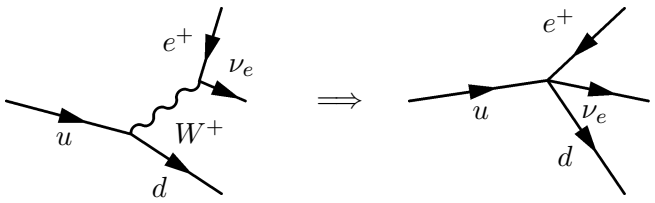
- $d[\mathcal{O}] > 4$: irrelevant perturbation. The running of the coupling constants takes the theory back to the fixed point.
- $d[\mathcal{O}] < 4$: relevant perturbation. At low energies the theory flows away from the scale-invariant theory.
- $d[\mathcal{O}] = 4$: marginal deformation. The direction of the flow cannot be decided only on dimensional grounds.

As an example, let us consider first a massless fermion theory perturbed by a four-fermion interaction term

$$\mathcal{L} = i\overline{\psi}\partial\!\!\!/\psi - \frac{1}{M^2}(\overline{\psi}\psi)^2. \tag{408}$$

This is indeed a perturbation by an irrelevant operator, since in four-dimensions $[\psi] = \frac{3}{2}$. Interactions generated by the extra term are suppressed at low energies since typically their effects are weighted by the dimensionless factor $\frac{E^2}{M^2}$, where E is the energy scale of the process. This means that as we try to capture the relevant physics at lower and lower energies the effect of the perturbation is weaker and weaker rendering in the infrared limit $E \to 0$ again a free theory. Hence, the irrelevant perturbation in (408) makes the theory flow back to the fixed point.

On the other hand relevant operators dominate the physics at low energies. This is the case, for example, of a mass term. As we lower the energy the mass becomes more important and once the energy goes below the mass of the field its dynamics is completely dominated by the mass term. This is, for example, how Fermi's theory of weak interactions emerges from the standard model at energies below the mass of the $W^{\pm}$ boson



At energies below $M_W = 80.4$ GeV the dynamics of the W^+ boson is dominated by its mass term and therefore becomes nonpropagating, giving rise to the effective four-fermion Fermi theory.

To summarize our discussion so far, we found that while relevant operators dominate the dynamics in the infrared, taking the theory away from the fixed point, irrelevant perturbations become suppressed in the same limit. Finally we consider the effect of marginal operators. As an example we take the interaction term in massless QED, $\mathcal{O} = \overline{\psi}\gamma^\mu\psi\, A_\mu$. Taking into account that in $d = 4$ the dimension of the electromagnetic potential is $[A_\mu] = 1$ the operator $\mathcal{O}$ is a marginal perturbation. In order to decide whether the fixed point theory

$$\mathcal{L}_0 = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\overline{\psi}\not{D}\,\psi \tag{409}$$

is restored at low energies or not we need to study the perturbed theory in more detail. This we have done in section 8.1 where we learned that the effective coupling in QED decreases at low energies. Then we conclude that the perturbed theory flows towards the fixed point in the infrared.

As an example of a marginal operator with the opposite behavior we can write the Lagrangian for a SU(N_c) gauge theory, $\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\,\mu\nu}$, as

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}\left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu\right)\left(\partial^\mu A^{a\,\nu} - \partial^\nu A^{a\,\mu}\right) - 4g f^{abc} A^a_\mu A^b_\nu\, \partial^\mu A^{c\,\nu} \\ &+ g^2 f^{abc} f^{ade} A^b_\mu A^c_\nu A^{d\,\mu} A^{e\,\nu} \equiv \mathcal{L}_0 + \mathcal{O}_g, \end{aligned} \tag{410}$$

i.e. a marginal perturbation of the free theory described by $\mathcal{L}_0$, which is obviously a fixed point under renormalization group transformations. Unlike the case of QED we know that the full theory is asymptotically free, so the coupling constant grows at low energies. This implies that the operator $\mathcal{O}_g$ becomes more and more important in the infrared and therefore the theory flows away the fixed point in this limit.

It is very important to notice here that in the Wilsonian view the cutoff is not necessarily regarded as just some artifact to remove infinities but actually has a physical origin. For example in the case of Fermi's theory of β-decay there is a natural cutoff $\Lambda = M_W$ at which the theory has to be replaced by the standard model. In the case of the standard model itself the cutoff can be taken at Planck scale $\Lambda \simeq 10^{19}$ GeV or the Grand Unification scale $\Lambda \simeq 10^{16}$ GeV, where new degrees of freedom are expected to become relevant. The cutoff serves the purpose of cloaking the range of energies at which new physics has to be taken into account.

Provided that in the Wilsonian approach the quantum theory is always defined with a physical cutoff, there is no fundamental difference between renormalizable and nonrenormalizable theories. Actually, a renormalizable field theory, like the standard model, can generate nonrenormalizable operators at low energies such as the effective four-fermion interaction of Fermi's theory. They are not sources of any trouble if we are interested in the physics at scales much below the cutoff, $E \ll \Lambda$, since their contribution to the amplitudes will be suppressed by powers of $\frac{E}{\Lambda}$.

9 Special topics

9.1 Creation of particles by classical fields

Particle creation by a classical source. In a free quantum field theory the total number of particles contained in a given state of the field is a conserved quantity. For example, in the case of the quantum scalar field studied in section 3 we have that the number operator commutes with the Hamiltonian

$$\widehat{n} \equiv \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_k}\alpha^\dagger(\vec{k})\alpha(\vec{k}), \qquad [\widehat{H}, \widehat{n}] = 0. \tag{411}$$

This means that any states with a well-defined number of particle excitations will preserve this number at all times. The situation, however, changes as soon as interactions are introduced, since in this case particles can be created and/or destroyed as a result of the dynamics.

Another case in which the number of particles might change is if the quantum theory is coupled to a classical source. The archetypical example of such a situation is the Schwinger effect, in which a classical strong electric field produces the creation of electron-positron pairs out of the vacuum. However, before plunging into this more involved situation we can illustrate the relevant physics involved in the creation of particles by classical sources with the help of the simplest example: a free scalar field theory coupled to a classical external source $J(x)$. The action for such a theory can be written as

$$S = \int d^4x \left[\frac{1}{2}\partial_\mu\phi(x)\partial^\mu\phi(x) - \frac{m^2}{2}\phi(x)^2 + J(x)\phi(x)\right], \tag{412}$$

where $J(x)$ is a real function of the coordinates. Its identification with a classical source is obvious once we calculate the equations of motion

$$\left(\nabla^2 + m^2\right)\phi(x) = J(x). \tag{413}$$

Our plan is to quantize this theory but, unlike the case analyzed in section 3, now the presence of the source $J(x)$ makes the situation a bit more involved. The general solution to the equations of motion can be written in terms of the retarded Green function for the Klein-Gordon equation as

$$\phi(x) = \phi_0(x) + i\int d^4x'\, G_R(x-x')J(x'), \tag{414}$$

where $\phi_0(x)$ is a general solution to the homogeneous equation and

$$\begin{aligned} G_R(t,\vec{x}) &= \int \frac{d^4k}{(2\pi)^4}\frac{i}{k^2-m^2+i\epsilon\,\mathrm{sign}(k^0)}e^{-ik\cdot x} \\ &= i\,\theta(t)\int\frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_k}\left(e^{-i\omega_k t+\vec{k}\cdot\vec{x}} - e^{i\omega_k t - i\vec{p}\cdot\vec{x}}\right), \end{aligned} \tag{415}$$

with $\theta(x)$ the Heaviside step function. The integration contour to evaluate the integral over p^0 surrounds the poles at $p^0 = \pm\omega_k$ from above. Since $G_R(t,\vec{x}) = 0$ for $t < 0$, the function $\phi_0(x)$ corresponds to the solution of the field equation at $t \to -\infty$, before the interaction with the external source[21]

To make the argument simpler we assume that $J(x)$ is switched on at $t = 0$, and only last for a time τ, that is

$$J(t,\vec{x}) = 0 \qquad \text{if } t < 0 \text{ or } t > \tau. \tag{416}$$

We are interested in a solution of (413) for times after the external source has been switched off, $t > \tau$. In this case the expression (415) can be written in terms of the Fourier modes $\widetilde{J}(\omega,\vec{k})$ of the source as

$$\phi(t,\vec{x}) = \phi_0(x) + i\int\frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_k}\left[\widetilde{J}(\omega_k,\vec{k})e^{-i\omega_k t+i\vec{k}\cdot\vec{x}} - \widetilde{J}(\omega_k,\vec{k})^* e^{i\omega_k t - i\vec{k}\cdot\vec{x}}\right]. \tag{417}$$

On the other hand, the general solution $\phi_0(x)$ has been already computed in Eq. (77). Combining this result with Eq. (417) we find the following expression for the late time general solution to the Klein-Gordon equation in the presence of the source

$$\begin{aligned} \phi(t,x) &= \int\frac{d^3k}{(2\pi)^3}\frac{1}{\sqrt{2\omega_k}}\left\{\left[\alpha(\vec{k}) + \frac{i}{\sqrt{2\omega_k}}\widetilde{J}(\omega_k,\vec{k})\right]e^{-i\omega_k t+i\vec{k}\cdot\vec{x}}\right. \\ &+ \left.\left[\alpha^*(\vec{k}) - \frac{i}{\sqrt{2\omega_k}}\widetilde{J}(\omega_k,\vec{k})^*\right]e^{i\omega_k t - i\vec{k}\cdot\vec{x}}\right\}. \end{aligned} \tag{418}$$

We should not forget that this is a solution valid for times $t > \tau$, i.e. once the external source has been disconnected. On the other hand, for $t < 0$ we find from Eqs. (414) and (415) that the general solution is given by Eq. (77).

Now we can proceed to quantize the theory. The conjugate momentum $\pi(x) = \partial_0\phi(x)$ can be computed from Eqs. (77) and (418). Imposing the canonical equal time commutation relations (74) we find that $\alpha(\vec{k})$, $\alpha^\dagger(\vec{k})$ satisfy the creation-annihilation algebra (51). From our previous calculation we find that for $t > \tau$ the expansion of the operator $\phi(x)$ in terms of the creation-annihilation operators $\alpha(\vec{k})$, $\alpha^\dagger(\vec{k})$ can be obtained from the one for $t < 0$ by the replacement

$$\alpha(\vec{k}) \quad\longrightarrow\quad \beta(\vec{k}) \equiv \alpha(\vec{k}) + \frac{i}{\sqrt{2\omega_k}}\widetilde{J}(\omega_k,\vec{k}),$$

[21] We could have taken instead the advanced propagator $G_A(x)$ in which case $\phi_0(x)$ would correspond to the solution to the equation at large times, after the interaction with $J(x)$.

$$\alpha^\dagger(\vec{k}) \quad \longrightarrow \quad \beta^\dagger(\vec{k}) \equiv \alpha^\dagger(\vec{k}) - \frac{i}{\sqrt{2\omega_k}} \widetilde{J}(\omega_k, \vec{k})^*. \tag{419}$$

Actually, since $\widetilde{J}(\omega_k, \vec{k})$ is a c-number, the operators $\beta(\vec{k})$, $\beta^\dagger(\vec{k})$ satisfy the same algebra as $\alpha(\vec{k})$, $\alpha^\dagger(\vec{k})$ and therefore can be interpreted as well as a set of creation-annihilation operators. This means that we can define two vacuum states, $|0_-\rangle$, $|0_+\rangle$ associated with both sets of operators

$$\left.\begin{array}{l} \alpha(\vec{k})|0_-\rangle = 0 \\ \\ \beta(\vec{k})|0_+\rangle = 0 \end{array}\right\} \qquad \forall\ \vec{k}. \tag{420}$$

For an observer at $t < 0$, $\alpha(\vec{k})$ and $\alpha(\vec{k})$ are the natural set of creation-annihilation operators in terms of which to expand the field operator $\phi(x)$. After the usual zero-point energy subtraction the Hamiltonian is given by

$$\widehat{H}^{(-)} = \frac{1}{2} \int \frac{d^3k}{(2\pi)^3}\, \alpha^\dagger(\vec{k})\alpha(\vec{k}) \tag{421}$$

and the ground state of the spectrum for this observer is the vacuum $|0_-\rangle$. At the same time, a second observer at $t > \tau$ will also see a free scalar quantum field (the source has been switched off at $t = \tau$) and consequently will expand ϕ in terms of the second set of creation-annihilation operators $\beta(\vec{k})$, $\beta^\dagger(\vec{k})$. In terms of this operators the Hamiltonian is written as

$$\widehat{H}^{(+)} = \frac{1}{2} \int \frac{d^3k}{(2\pi)^3}\, \beta^\dagger(\vec{k})\beta(\vec{k}). \tag{422}$$

Then for this late-time observer the ground state of the Hamiltonian is the second vacuum state $|0_+\rangle$.

In our analysis we have been working in the Heisenberg picture, where states are time-independent and the time dependence comes in the operators. Therefore the states of the theory are globally defined. Suppose now that the system is in the "in" ground state $|0_-\rangle$. An observer at $t < 0$ will find that there are no particles

$$\widehat{n}^{(-)}|0_-\rangle = 0. \tag{423}$$

However the late-time observer will find that the state $|0_-\rangle$ contains an average number of particles given by

$$\langle 0_-|\widehat{n}^{(+)}|0_-\rangle = \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left|\widetilde{J}(\omega_k, \vec{k})\right|^2. \tag{424}$$

Moreover, $|0_-\rangle$ is no longer the ground state for the "out" observer. On the contrary, this state have a vacuum expectation value for $\widehat{H}^{(+)}$

$$\langle 0_-|\widehat{H}^{(+)}|0_-\rangle = \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} \left|\tilde{J}(\omega_k, \vec{k})\right|^2. \tag{425}$$

The key to understand what is going on here lies in the fact that the external source breaks the invariance of the theory under space-time translations. In the particular case we have studied here where $J(x)$ has support over a finite time interval $0 < t < \tau$, this implies that the vacuum is not invariant under time translations, so observers at different times will make different choices of vacuum that will not necessarily agree with each other. This is clear in our example. An observer in $t < \tau$ will choose the vacuum to be the lowest energy state of her Hamiltonian, $|0_-\rangle$. On the other hand, the second observer at late times $t > \tau$ will naturally choose $|0_+\rangle$ as the vacuum. However, for this second observer, the

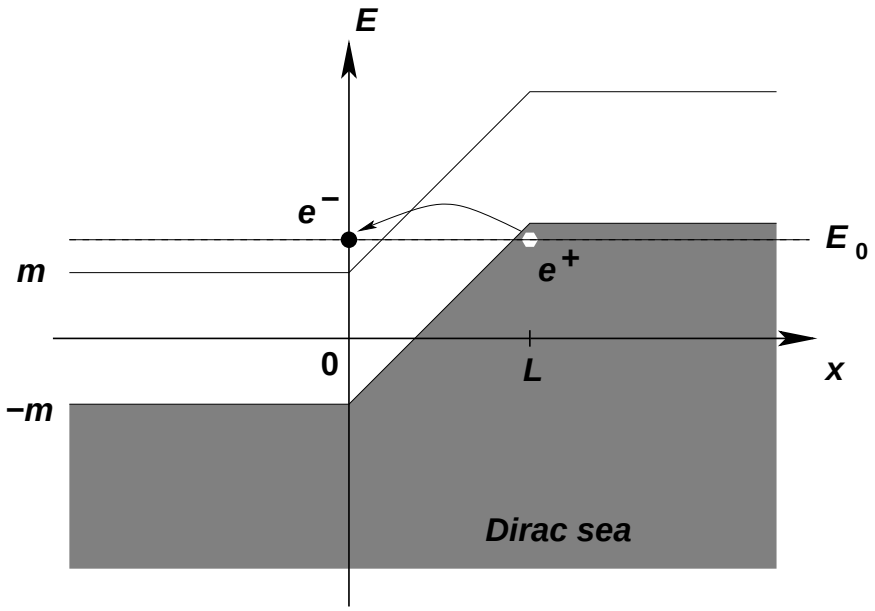


Fig. 18: Pair creation by a electric field in the Dirac sea picture.

state $|0_-\rangle$ is not the vacuum of his Hamiltonian, but actually an excited state that is a superposition of states with well-defined number of particles. In this sense it can be said that the external source has the effect of creating particles out of the "in" vacuum. Besides, this breaking of time translation invariance produces a violation in the energy conservation as we see from Eq. (425). Particles are actually created from the energy pumped into the system by the external source.

The Schwinger effect. A classical example of creation of particles by a external field was pointed out by Schwinger [44] and consists of the creation of electron-positron pairs by a strong electric field. In order to illustrate this effect we are going to follow a heuristic argument based on the Dirac sea picture and the WKB approximation.

In the absence of an electric field the vacuum state of a spin-$\frac{1}{2}$ field is constructed by filling all the negative energy states as depicted in Fig. 2. Let us now connect a constant electric field $\vec{\mathcal{E}} = \mathcal{E}\vec{u}_x$ in the range $0 < x < L$ created by a electrostatic potential

$$V(\vec{r}) = \begin{cases} 0 & x < 0 \\ -\mathcal{E}x & 0 < x < L \\ -\mathcal{E}L & x > L \end{cases} \tag{426}$$

After the field has been switched on, the Dirac sea looks like in Fig. 18. In particular we find that if $e\mathcal{E}L > 2m$ there are negative energy states at $x > L$ with the same energy as the positive energy states in the region $x < 0$. Therefore it is possible for an electron filling a negative energy state with energy close to $-2m$ to tunnel through the forbidden region into a positive energy state. The interpretation of such a process is the production of an electron-positron pair out of the electric field.

We can compute the rate at which such pairs are produced by using the WKB approximation. Focusing for simplicity on an electron on top of the Fermi surface near $x = L$ with energy E_0, the transmission coefficient in this approximation is given by[22]

$$T_{\rm WKB} \;=\; \exp\left[-2\int_{\frac{1}{e\mathcal{E}}\left(E_0-\sqrt{m^2+\vec{p}_T^{\,2}}\right)}^{\frac{1}{e\mathcal{E}}\left(E_0+\sqrt{m^2+\vec{p}_T^{\,2}}\right)} dx\,\sqrt{m^2-\left[E_0-e\mathcal{E}(x-x_0)\right]^2+\vec{p}_T^{\,2}}\right]$$

[22] Notice that the electron satisfy the relativistic dispersion relation $E = \sqrt{\vec{p}^{\,2}+m^2}+V$ and therefore $-p_x^2 = m^2-(E-V)^2+\vec{p}_T^{\,2}$. The integration limits are set by those values of x at which $p_x = 0$.

$$= \quad \exp\left[-\frac{\pi}{e\mathcal{E}}\left(\vec{p}_T^{\,2}+m^2\right)\right], \tag{427}$$

where $p_T^2 \equiv p_y^2 + p_z^2$. This gives the transition probability per unit time and per unit cross section $dydz$ for an electron in the Dirac sea with transverse momentum $\vec{p}_T$ and energy E_0. To get the total probability per unit time and per unit volume we have to integrate over all possible values of $\vec{p}_T$ and E_0. Actually, in the case of the energy, because of the relation between E_0 and the coordinate x at which the particle penetrates into the barrier we can write $\frac{dE_0}{2\pi} = \frac{e\mathcal{E}}{2\pi}dx$ and the total probability per unit time and per unit volume for the creation of a pair is given by

$$W = 2\left(\frac{e\mathcal{E}}{2\pi}\right)\int \frac{d^2p_T}{(2\pi)^2} e^{-\frac{\pi}{e\mathcal{E}}\left(\vec{p}_T^{\,2}+m^2\right)} = \frac{e^2\mathcal{E}^2}{4\pi^3} e^{-\frac{\pi\, m^2}{e\mathcal{E}}}, \tag{428}$$

where the factor of 2 accounts for the two polarizations of the electron.

Then production of electron-positron pairs is exponentially suppressed and it is only sizeable for strong electric fields. To estimate its order of magnitude it is useful to restore the powers of c and $\hbar$ in (428)

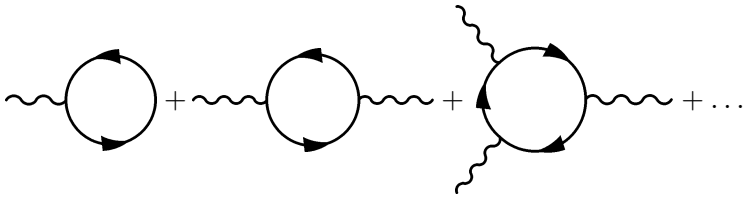
$$W = \frac{e^2\mathcal{E}^2}{4\pi^3 c\hbar^2} e^{-\frac{\pi\, m^2 c^3}{\hbar e\mathcal{E}}} \tag{429}$$

The exponential suppression of the pair production disappears when the electric field reaches the critical value $\mathcal{E}_{\rm crit}$ at which the exponent is of order one

$$\mathcal{E}_{\rm crit} = \frac{m^2c^3}{\hbar e} \simeq 1.3\times 10^{16}\,{\rm V\,cm^{-1}}. \tag{430}$$

This is indeed a very strong field which is extremely difficult to produce. A similar effect, however, takes place also in a time-varying electric field [45] and there is the hope that pair production could be observed in the presence of the alternating electric field produced by a laser.

The heuristic derivation that we followed here can be made more precise in QED. There the decay of the vacuum into electron-positron pairs can be computed from the imaginary part of the effective action $\Gamma[A_\mu]$ in the presence of a classical gauge potential A_μ

$i\Gamma[A_\mu] \quad \equiv$  $+\ldots$

$$= \quad \log\ \det\left[1 - ie\not{A}\frac{1}{i\not{\partial} - m}\right]. \tag{431}$$

This determinant can be computed using the standard heat kernel techniques. The probability of pair production is proportional to the imaginary part of $i\Gamma[A_\mu]$ and gives

$$W = \frac{e^2\mathcal{E}^2}{4\pi^3}\sum_{n=1}^{\infty}\frac{1}{n^2}e^{-n\frac{\pi\, m^2}{e\mathcal{E}}}. \tag{432}$$

Our simple argument based on tunneling in the Dirac sea gave only the leading term of Schwinger's result (432). The remaining terms can be also captured in the WKB approximation by taking into account the probability of production of several pairs, i.e. the tunneling of more than one electron through the barrier.

Here we have illustrated the creation of particles by semiclassical sources in quantum field theory using simple examples. Nevertheless, what we learned has important applications to the study of quantum fields in curved backgrounds. In quantum field theory in Minkowski space-time the vacuum state

is invariant under the Poincaré group and this, together with the covariance of the theory under Lorentz transformations, implies that all inertial observers agree on the number of particles contained in a quantum state. The breaking of such invariance, as happened in the case of coupling to a time-varying source analyzed above, implies that it is not possible anymore to define a state which would be recognized as the vacuum by all observers.

This is precisely the situation when fields are quantized on curved backgrounds. In particular, if the background is time-dependent (as it happens in a cosmological setup or for a collapsing star) different observers will identify different vacuum states. As a consequence what one observer call the vacuum will be full of particles for a different observer. This is precisely what is behind the phenomenon of Hawking radiation [46]. The emission of particles by a physical black hole formed from gravitational collapse of a star is the consequence of the fact that the vacuum state in the asymptotic past contain particles for an observer in the asymptotic future. As a consequence, a detector located far away from the black hole detects a stream of thermal radiation with temperature

$$T_{\rm Hawking} = \frac{\hbar c^3}{8\pi G_N\, k\, M} \tag{433}$$

where M is the mass of the black hole, G_N is Newton's constant and k is Boltzmann's constant. There are several ways in which this results can be obtained. A more heuristic way is perhaps to think of this particle creation as resulting from quantum tunneling of particles across the potential barrier posed by gravity [47].

9.2 Supersymmetry

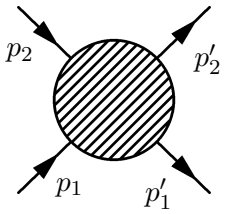
One of the things that we have learned in our journey around the landscape of quantum field theory is that our knowledge of the fundamental interactions in Nature is based on the idea of symmetry, and in particular gauge symmetry. The Lagrangian of the standard model can be written just including all possible renormalizable terms (i.e. with canonical dimension smaller o equal to 4) compatible with the gauge symmetry SU(3)×SU(2)×U(1) and Poincaré invariance. All attempts to go beyond start with the question of how to extend the symmetries of the standard model.

As explained in Section 5.1, in a quantum field theoretical description of the interaction of elementary particles the basic observable quantity to compute is the scattering or S-matrix giving the probability amplitude for the scattering of a number of incoming particles with a certain momentum into some final products

$$\mathcal{A}(\text{in} \longrightarrow \text{out}) = \langle \vec{p_1}', \ldots; \text{out} | \vec{p}_1, \ldots; \text{in} \rangle. \tag{434}$$

An explicit symmetry of the theory has to be necessarily a symmetry of the S-matrix. Hence it is fair to ask what is the largest symmetry of the S-matrix.

Let us ask this question in the simple case of the scattering of two particles with four-momenta p_1 and p_2 in the t-channel



We will make the usual assumptions regarding positivity of the energy and analyticity. Invariance of the theory under the Poincaré group implies that the amplitude can only depend on the scattering angle ϑ through

$$t = (p_1' - p_1)^2 = 2\left(m_1^2 - p_1 \cdot p_1'\right) = 2\left(m_1^2 - E_1 E_1' + |\vec{p}_1||\vec{p}_1{}'|\cos\vartheta\right). \tag{435}$$

If there would be any extra bosonic symmetry of the theory it would restrict the scattering angle to a set of discrete values. In this case the S-matrix cannot be analytic since it would vanish everywhere except for the discrete values selected by the extra symmetry.

Actually, the only way to extend the symmetry of the theory without renouncing to the analyticity of the scattering amplitudes is to introduce "fermionic" symmetries, i.e. symmetries whose generators are anticommuting objects [48]. This means that in addition to the generators of the Poincaré group[23] P^μ, $M^{\mu\nu}$ and the ones for the internal gauge symmetries G, we can introduce a number of fermionic generators Q^I_a, $\overline{Q}_{\dot{a}\,I}$ ($I = 1, \ldots, \mathcal{N}$), where $\overline{Q}_{\dot{a}\,I} = (Q^I_a)^\dagger$. The most general algebra that these generators satisfy is the $\mathcal{N}$-extended supersymmetry algebra [49]

$$\begin{aligned} \{Q^I_a, \overline{Q}_{\dot{b}\,J}\} &= 2\sigma^\mu_{a\dot{b}} P_\mu \delta^I{}_J, \\ \{Q^I_a, Q^J_b\} &= 2\varepsilon_{ab}\mathcal{Z}^{IJ}, \qquad (436) \\ \{\overline{Q}^I_{\dot{a}}, \overline{Q}^J_{\dot{b}}\} &= 2\varepsilon_{\dot{a}\dot{b}}\overline{\mathcal{Z}}^{IJ}, \qquad (437) \end{aligned}$$

where $\mathcal{Z}^{IJ} \in \mathbb{C}$ commute with any other generator and satisfies $\mathcal{Z}^{IJ} = -\mathcal{Z}^{JI}$. Besides we have the commutators that determine the Poincaré transformations of the fermionic generators Q^I_a, $Q_{\dot{a}\,J}$

$$\begin{aligned} [Q^I_a, P^\mu] &= [\overline{Q}_{\dot{a}\,I}, P^\mu] = 0, \\ [Q^I_a, M^{\mu\nu}] &= \frac{1}{2}(\sigma^{\mu\nu})_a{}^b Q^I_b, \qquad (438) \\ [\overline{Q}_{a\,I}, M^{\mu\nu}] &= -\frac{1}{2}(\overline{\sigma}^{\mu\nu})_{\dot{a}}{}^{\dot{b}} \overline{Q}_{\dot{b}\,I}, \end{aligned}$$

where $\sigma^{0i} = -i\sigma^i$, $\sigma^{ij} = \varepsilon^{ijk}\sigma^k$ and $\overline{\sigma}^{\mu\nu} = (\sigma^{\mu\nu})^\dagger$. These identities simply mean that Q^I_a, $\overline{Q}_{\dot{a}\,J}$ transform respectively in the $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$ and $(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$ representations of the Lorentz group.

We know that the presence of a global symmetry in a theory implies that the spectrum can be classified in multiplets with respect to that symmetry. In the case of supersymmetry start with the case case $\mathcal{N} = 1$ in which there is a single pair of supercharges Q_a, $\overline{Q}_{\dot{a}}$ satisfying the algebra

$$\{Q_a, \overline{Q}_{\dot{b}}\} = 2\sigma^\mu_{a\dot{b}} P_\mu, \qquad \{Q_a, Q_b\} = \{\overline{Q}_{\dot{a}}, \overline{Q}_{\dot{b}}\} = 0. \qquad (439)$$

Notice that in the $\mathcal{N} = 1$ case there is no possibility of having central charges.

We study now the representations of the supersymmetry algebra (439), starting with the massless case. Given a state $|k\rangle$ satisfying $k^2 = 0$, we can always find a reference frame where the four-vector k^μ takes the form $k^\mu = (E, 0, 0, E)$. Since the theory is Lorentz covariant we can obtain the representation of the supersymmetry algebra in this frame where the expressions are simpler. In particular, the right-hand side of the first anticommutator in Eq. (439) is given by

$$2\sigma^\mu_{a\dot{b}} P_\mu = 2(P^0 - \sigma^3 P^3) = \begin{pmatrix} 0 & 0 \\ 0 & 4E \end{pmatrix}. \qquad (440)$$

Therefore the algebra of supercharges in the massless case reduces to

$$\begin{aligned} \{Q_1, Q_1^\dagger\} &= \{Q_1, Q_2^\dagger\} = 0, \\ \{Q_2, Q_2^\dagger\} &= 4E. \qquad (441) \end{aligned}$$

The commutator $\{Q_1, Q_1^\dagger\} = 0$ implies that the action of Q_1 on any state gives a zero-norm state of the Hilbert space $\|Q_1|\Psi\rangle\| = 0$. If we want the theory to preserve unitarity we must eliminate these null

[23] The generators $M^{\mu\nu}$ are related with the ones for boost and rotations introduced in section 4.1 by $J^i \equiv M^{0i}$, $M^i = \frac{1}{2}\varepsilon^{ijk}M^{jk}$. In this section we also use the "dotted spinor" notation, in which spinors in the $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$ and $(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$ representations of the Lorentz group are indicated respectively by undotted $(a, b, \ldots)$ and dotted $(\dot{a}, \dot{b}, \ldots)$ indices.

states from the spectrum. This is equivalent to setting $Q_1 \equiv 0$. On the other hand, in terms of the second generator Q_2 we can define the operators

$$a = \frac{1}{2\sqrt{E}} Q_2, \qquad a^\dagger = \frac{1}{2\sqrt{E}} Q_2^\dagger, \tag{442}$$

which satisfy the algebra of a pair of fermionic creation-annihilation operators, $\{a, a^\dagger\} = 1$, $a^2 = (a^\dagger)^2 = 0$. Starting with a vacuum state $a|\lambda\rangle = 0$ with helicity λ we can build the massless multiplet

$$|\lambda\rangle, \qquad |\lambda + \tfrac{1}{2}\rangle \equiv a^\dagger |\lambda\rangle. \tag{443}$$

Here we consider two important cases:

– Scalar multiplet: we take the vacuum state to have zero helicity $|0^+\rangle$ so the multiplet consists of a scalar and a helicity-$\frac{1}{2}$ state

$$|0^+\rangle, \qquad |\tfrac{1}{2}\rangle \equiv a^\dagger |0^+\rangle. \tag{444}$$

However, this multiplet is not invariant under the CPT transformation which reverses the sign of the helicity of the states. In order to have a CPT-invariant theory we have to add to this multiplet its CPT-conjugate which can be obtain from a vacuum state with helicity $\lambda = -\frac{1}{2}$

$$|0^-\rangle, \qquad |-\tfrac{1}{2}\rangle. \tag{445}$$

Putting them together we can combine the two zero helicity states with the two fermionic ones into the degrees of freedom of a complex scalar field and a Weyl (or Majorana) spinor.

– Vector multiplet: now we take the vacuum state to have helicity $\lambda = \frac{1}{2}$, so the multiplet contains also a massless state with helicity $\lambda = 1$

$$|\tfrac{1}{2}\rangle, \qquad |1\rangle \equiv a^\dagger |\tfrac{1}{2}\rangle. \tag{446}$$

As with the scalar multiplet we add the CPT conjugated obtained from a vacuum state with helicity $\lambda = -1$

$$|-\tfrac{1}{2}\rangle, \qquad |-1\rangle, \tag{447}$$

which together with (446) give the propagating states of a gauge field and a spin-$\frac{1}{2}$ gaugino.

In both cases we see the trademark of supersymmetric theories: the number of bosonic and fermionic states within a multiplet are the same.

In the case of extended supersymmetry we have to repeat the previous analysis for each supersymmetry charge. At the end, we have $\mathcal{N}$ sets of fermionic creation-annihilation operators $\{a^I, a_I^\dagger\} = \delta^I{}_J$, $(a_I)^2 = (a_I^\dagger)^2 = 0$. Let us work out the case of $\mathcal{N} = 8$ supersymmetry. Since for several reasons we do not want to have states with helicity larger than 2, we start with a vacuum state $|-2\rangle$ of helicity $\lambda = -2$. The rest of the states of the supermultiplet are obtained by applying the eight different creation operators $a_I^\dagger$ to the vacuum:

$$\begin{aligned}
\lambda = 2: &\quad a_1^\dagger \dots a_8^\dagger |-2\rangle && \binom{8}{8} = 1 \text{ state}, \\
\lambda = \frac{3}{2}: &\quad a_{I_1}^\dagger \dots a_{I_7}^\dagger |-2\rangle && \binom{8}{7} = 8 \text{ states}, \\
\lambda = 1: &\quad a_{I_1}^\dagger \dots a_{I_6}^\dagger |-2\rangle && \binom{8}{6} = 28 \text{ states},
\end{aligned}$$

$$\begin{aligned}
\lambda = \frac{1}{2}: &\quad a^\dagger_{I_1} \dots a^\dagger_{I_5} |-2\rangle && \binom{8}{5} = 56 \text{ states}, \\
\lambda = 0: &\quad a^\dagger_{I_1} \dots a^\dagger_{I_4} |-2\rangle && \binom{8}{4} = 70 \text{ states}, \\
\lambda = -\frac{1}{2}: &\quad a^\dagger_{I_1} a^\dagger_{I_2} a^\dagger_{I_3} |-2\rangle && \binom{8}{3} = 56 \text{ states}, \\
\lambda = -1: &\quad a^\dagger_{I_1} a^\dagger_{I_2} |-2\rangle && \binom{8}{2} = 28 \text{ states}, \\
\lambda = -\frac{3}{2}: &\quad a^\dagger_{I_1} |-2\rangle && \binom{8}{1} = 8 \text{ states}, \\
\lambda = -2: &\quad |-2\rangle && 1 \text{ state}.
\end{aligned} \tag{448}$$

Putting together the states with opposite helicity we find that the theory contains:

- 1 spin-2 field $g_{\mu\nu}$ (a graviton),
- 8 spin-$\frac{3}{2}$ gravitino fields ψ^I_μ,
- 28 gauge fields $A^{[IJ]}_\mu$,
- 56 spin-$\frac{1}{2}$ fermions $\psi^{[IJK]}$,
- 70 scalars $\phi^{[IJKL]}$,

where by $[IJ...]$ we have denoted that the indices are antisymmetrized. We see that, unlike the massless multiplets of $\mathcal{N} = 1$ supersymmetry studied above, this multiplet is CPT invariant by itself. As in the case of the massless $\mathcal{N} = 1$ multiplet, here we also find as many bosonic as fermionic states:

$$\begin{aligned}
\text{bosons:} \quad & 1 + 28 + 70 + 28 + 1 = 128 \quad \text{states}, \\
\text{fermions:} \quad & 8 + 56 + 56 + 8 = 128 \quad \text{states}.
\end{aligned}$$

Now we study briefly the case of massive representations $|k\rangle$, $k^2 = M^2$. Things become simpler if we work in the rest frame where $P^0 = M$ and the spatial components of the momentum vanish. Then, the supersymmetry algebra becomes:

$$\{Q^I_a, \overline{Q}_{\dot{b}\,J}\} = 2M\delta_{a\dot{b}}\delta^I{}_J. \tag{449}$$

We proceed now in a similar way to the massless case by defining the operators

$$a^I_a \equiv \frac{1}{\sqrt{2M}} Q^I_a, \qquad a^\dagger_{\dot{a}\,I} \equiv \frac{1}{\sqrt{2M}} \overline{Q}_{\dot{a}\,I}. \tag{450}$$

The multiplets are found by choosing a vacuum state with a definite spin. For example, for $\mathcal{N} = 1$ and taking a spin-0 vacuum $|0\rangle$ we find three states in the multiplet transforming irreducibly with respect to the Lorentz group:

$$|0\rangle, \qquad a^\dagger_{\dot{a}}|0\rangle, \qquad \varepsilon^{\dot{a}\dot{b}} a^\dagger_{\dot{a}} a^\dagger_{\dot{b}} |0\rangle, \tag{451}$$

which, once transformed back from the rest frame, correspond to the physical states of two spin-0 bosons and one spin-$\frac{1}{2}$ fermion. For $\mathcal{N}$-extended supersymmetry the corresponding multiplets can be worked out in a similar way.

The equality between bosonic and fermionic degrees of freedom is at the root of many of the interesting properties of supersymmetric theories. For example, in section 4 we computed the divergent vacuum energy contributions for each real bosonic or fermionic propagating degree of freedom is[24]

$$E_{\text{vac}} = \pm \frac{1}{2} \delta(\vec{0}) \int d^3p\, \omega_p, \tag{452}$$

[24]For a boson, this can be read off Eq. (80). In the case of fermions, the result of Eq. (134) gives the vacuum energy contribution of the four real propagating degrees of freedom of a Dirac spinor.

where the sign $\pm$ corresponds respectively to bosons and fermions. Hence, for a supersymmetric theory the vacuum energy contribution exactly cancels between bosons and fermions. This boson-fermion degeneracy is also responsible for supersymmetric quantum field theories being less divergent than non-supersymmetric ones.

Appendix: A crash course in Group Theory

In this Appendix we summarize some basic facts about Group Theory. Given a group G a representation of G is a correspondence between the elements of G and the set of linear operators acting on a vector space V, such that for each element of the group $g \in G$ there is a linear operator $D(g)$

$$D(g) : V \longrightarrow V \tag{453}$$

satisfying the group operations

$$D(g_1)D(g_2) = D(g_1 g_2), \qquad D(g_1^{-1}) = D(g_1)^{-1}, \qquad g_1, g_2 \in \mathcal{G}. \tag{454}$$

The representation $D(g)$ is irreducible if and only if the only operators $A : V \to V$ commuting with all the elements of the representation $D(g)$ are the ones proportional to the identity

$$[D(g), A] = 0, \ \ \forall g \qquad \Longleftrightarrow \qquad A = \lambda \mathbf{1}, \quad \lambda \in \mathbb{C} \tag{455}$$

More intuitively, we can say that a representation is irreducible if there is no proper subspace $U \subset V$ (i.e. $U \neq V$ and $U \neq \emptyset$) such that $D(g)U \subset U$ for every element $g \in G$.

Here we are specially interested in Lie groups whose elements are labelled by a number of continuous parameters. In mathematical terms this means that a Lie group is a manifold $\mathcal{M}$ together with an operation $\mathcal{M} \times \mathcal{M} \longrightarrow \mathcal{M}$ that we will call multiplication that satisfies the associativity property $g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$ together with the existence of unity $g\mathbf{1} = \mathbf{1}g = g$,for every $g \in \mathcal{M}$ and inverse $gg^{-1} = g^{-1}g = \mathbf{1}$.

The simplest example of a Lie group is SO(2), the group of rotations in the plane. Each element $R(\theta)$ is labelled by the rotation angle θ, with the multiplication acting as $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$. Because the angle θ is defined only modulo 2π, the manifold of SO(2) is a circumference S^1.

One of the interesting properties of Lie groups is that in a neighborhood of the identity element they can be expressed in terms of a set of generators T^a ($a = 1, \ldots, \dim G$) as

$$D(g) = \exp(-i\alpha_a T^a) \equiv \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \alpha_{a_1} \ldots \alpha_{a_n} T^{a_1} \ldots T^{a_n}, \tag{456}$$

where $\alpha_a \in \mathbb{C}$ are a set of coordinates of $\mathcal{M}$ in a neighborhood of $\mathbf{1}$. Because of the general Baker-Campbell-Haussdorf formula, the multiplication of two group elements is encoded in the value of the commutator of two generators, that in general has the form

$$[T^a, T^b] = i f^{abc} T^c, \tag{457}$$

where $f^{abc} \in \mathbb{C}$ are called the structure constants. The set of generators with the commutator operation form the Lie algebra associated with the Lie group. Hence, given a representation of the Lie algebra of generators we can construct a representation of the group by exponentiation (at least locally near the identity).

We illustrate these concept with some particular examples. For SU(2) each group element is labelled by three real number α_i, $i = 1, 2, 3$. We have two basic representations: one is the fundamental representation (or spin $\frac{1}{2}$) defined by

$$D_{\frac{1}{2}}(\alpha_i) = e^{-\frac{i}{2}\alpha_i \sigma^i}, \tag{458}$$

with σ^i the Pauli matrices. The second one is the adjoint (or spin 1) representation which can be written as

$$D_1(\alpha_i) = e^{-i\alpha_i J^i}, \tag{459}$$

where

$$J^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \qquad J^2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad J^3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{460}$$

Actually, J^i ($i = 1, 2, 3$) generate rotations around the x, y and z axis respectively. Representations of spin $j \in \mathbb{N} + \frac{1}{2}$ can be also constructed with dimension

$$\dim D_j(g) = 2j + 1. \tag{461}$$

As a second example we consider SU(3). This group has two basic three-dimensional representations denoted by $\mathbf{3}$ and $\overline{\mathbf{3}}$ which in QCD are associated with the transformation of quarks and antiquarks under the color gauge symmetry SU(3). The elements of these representations can be written as

$$D_{\mathbf{3}}(\alpha^a) = e^{\frac{i}{2}\alpha^a\lambda_a}, \qquad D_{\overline{\mathbf{3}}}(\alpha^a) = e^{-\frac{i}{2}\alpha^a\lambda_a^T} \qquad (a = 1, \ldots, 8), \tag{462}$$

where λ_a are the eight hermitian Gell-Mann matrices

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
\lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \qquad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\
\lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \lambda_8 = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & -\frac{2}{\sqrt{3}} \end{pmatrix}.
\end{aligned} \tag{463}$$

Hence the generators of the representations $\mathbf{3}$ and $\overline{\mathbf{3}}$ are given by

$$T^a(\mathbf{3}) = \frac{1}{2}\lambda_a, \qquad T^a(\overline{\mathbf{3}}) = -\frac{1}{2}\lambda_a^T. \tag{464}$$

Irreducible representations can be classified in three groups: real, complex and pseudoreal.

– Real representations: a representation is said to be real if there is a *symmetric matrix* S which acts as intertwiner between the generators and their complex conjugates

$$\overline{T}^a = -ST^aS^{-1}, \qquad S^T = S. \tag{465}$$

This is for example the case of the adjoint representation of SU(2) generated by the matrices (460)

– Pseudoreal representations: are the ones for which an *antisymmetric matrix* S exists with the property

$$\overline{T}^a = -ST^aS^{-1}, \qquad S^T = -S. \tag{466}$$

As an example we can mention the spin-$\frac{1}{2}$ representation of SU(2) generated by $\frac{1}{2}\sigma^i$.

– Complex representations: finally, a representation is complex if the generators and their complex conjugate are not related by a similarity transformation. This is for instance the case of the two three-dimensional representations $\mathbf{3}$ and $\overline{\mathbf{3}}$ of SU(3).

There are a number of invariants that can be constructed associated with an irreducible representation R of a Lie group G and that can be used to label such a representation. If T_R^a are the generators in a certain representation R of the Lie algebra, it is easy to see that the matrix $\sum_{a=1}^{\dim G} T_R^a T_R^a$ commutes with every generator T_R^a. Therefore, because of Schur's lemma, it has to be proportional to the identity[25]. This defines the Casimir invariant $C_2(R)$ as

$$\sum_{a=1}^{\dim G} T_R^a T_R^a = C_2(R)\mathbf{1}. \tag{467}$$

A second invariant $T_2(R)$ associated with a representation R can also be defined by the identity

$$\mathrm{Tr}\, T_R^a T_R^b = T_2(R)\delta^{ab}. \tag{468}$$

Actually, taking the trace in Eq. (467) and combining the result with (468) we find that both invariants are related by the identity

$$C_2(R)\dim R = T_2(R)\dim G, \tag{469}$$

with $\dim R$ the dimension of the representation R.

These two invariants appear frequently in quantum field theory calculations with nonabelian gauge fields. For example $T_2(R)$ comes about as the coefficient of the one-loop calculation of the beta-function for a Yang-Mills theory with gauge group G. In the case of SU(N), for the fundamental representation, we find the values

$$C_2(\text{fund}) = \frac{N^2-1}{2N}, \qquad T_2(\text{fund}) = \frac{1}{2}, \tag{470}$$

whereas for the adjoint representation the results are

$$C_2(\text{adj}) = N, \qquad T_2(\text{adj}) = N. \tag{471}$$

A third invariant $A(R)$ is specially important in the calculation of anomalies. As discussed in section (7), the chiral anomaly in gauge theories is proportional to the group-theoretical factor $\mathrm{Tr}\,\left[T_R^a\{T_R^b, T_R^c\}\right]$ This leads us to define $A(R)$ as

$$\mathrm{Tr}\,\left[T_R^a\{T_R^b, T_R^c\}\right] = A(R)d^{abc}, \tag{472}$$

where d^{abc} is symmetric in its three indices and does not depend on the representation. Therefore, the cancellation of anomalies in a gauge theory with fermions transformed in the representation R of the gauge group is guaranteed if the corresponding invariant $A(R)$ vanishes.

It is not difficult to prove that $A(R) = 0$ if the representation R is either real or pseudoreal. Indeed, if this is the case, then there is a matrix S (symmetric or antisymmetric) that intertwins the generators T_R^a and their complex conjugates $\overline{T}_R^a = -ST_R^aS^{-1}$. Then, using the hermiticity of the generators we can write

$$\mathrm{Tr}\,\left[T_R^a\{T_R^b, T_R^c\}\right] = \mathrm{Tr}\,\left[T_R^a\{T_R^b, T_R^c\}\right]^T = \mathrm{Tr}\,\left[\overline{T}_R^a\{\overline{T}_R^b, \overline{T}_R^c\}\right]. \tag{473}$$

[25] Schur's lemma states that if there is a matrix A that commutes with all elements of an irreducible representation of a Lie algebra, then $A = \lambda\mathbf{1}$, for some $\lambda \in \mathbb{C}$.

Now, using (465) or (466) we have

$$\text{Tr}\,\left[\overline{T}^a_R\{\overline{T}^b_R,\overline{T}^c_R\}\right] = -\text{Tr}\,\left[ST^a_RS^{-1}\{ST^b_RS^{-1},ST^c_RS^{-1}\}\right] = -\text{Tr}\,\left[T^a_R\{T^b_R,T^c_R\}\right], \qquad (474)$$

which proves that Tr $\left[T^a_R\{T^b_R,T^c_R\}\right]$ and therefore $A(R)=0$ whenever the representation is real or pseudoreal. Since the gauge anomaly in four dimensions is proportional to $A(R)$ this means that anomalies appear only when the fermions transform in a complex representation of the gauge group.

References

[1] L. Álvarez-Gaumé and M. A. Vázquez-Mozo, *An Invitation to Quantum Field Theory*, Springer 2011.

[2] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields*, McGraw-Hill 1965.

[3] C. Itzykson and J.-B. Zuber, *Quantum Field Theory*, McGraw-Hill 1980.

[4] P. Ramond, *Field Theory: A Modern Primer*, Addison-Wesley 1990.

[5] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison Wesley 1995.

[6] S. Weinberg, *The Quantum Theory of Fields*, Vols. 1-3, Cambridge 1995

[7] P. Deligne et al. (editors), *Quantum Fields and Strings: a Course for Mathematicians*, American Mathematical Society 1999.

[8] A. Zee, *Quantum Field Theory in a Nutshell*, Princeton 2003.

[9] B. S. DeWitt, *The Global Approach to Quantum Field Theory*, Vols. 1 & 2, Oxford 2003.

[10] V. P. Nair, *Quantum Field Theory. A Modern Perspective*, Springer 2005.

[11] T. Banks, *Modern Quantum Field Theory*, Cambridge 2008.

[12] O. Klein, *Die Reflexion von Elektronen an einem Potentialsprung nach der Relativischen Dynamik von Dirac*, Z. Phys. **53** (1929) 157.

[13] B. R. Holstein, *Klein's paradox*, Am. J. Phys. **66** (1998) 507.

[14] N. Dombey and A. Calogeracos, *Seventy years of the Klein paradox*, Phys. Rept. **315** (1999) 41.
N. Dombey and A. Calogeracos, *History and Physics of the Klein Paradox*, Contemp. Phys. **40** (1999) 313 (`quant-ph/9905076`).

[15] F. Sauter, *Zum Kleinschen Paradoxon*, Z. Phys. **73** (1932) 547.

[16] H. B. G. Casimir, *On the attraction between two perfectly conducting plates*, Proc. Kon. Ned. Akad. Wet. **60** (1948) 793.

[17] G. Plunien, B. Müller and W. Greiner, *The Casimir Effect*, Phys. Rept. **134** (1986) 87.
K. A. Milton, *The Casimir Effect: Physical Manifestation of Zero-Point Energy*, (`hep-th/9901011`).
K. A. Milton, *The Casimir effect: recent controversies and progress*, J. Phys. **A37** (2004) R209 (`hep-th/0406024`).
S. K. Lamoreaux, *The Casimir force: background, experiments, and applications*, Rep. Prog. Phys. **68** (2005) 201.

[18] M. J. Sparnaay, *Measurement of attractive forces between flat plates*, Physica **24** (1958) 751.

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover 1972.

[20] Y. Aharonov and D. Bohm, *Significance of the electromagnetic potentials in the quantum theory*, Phys. Rev. **115** (1955) 485.

[21] P. A. M. Dirac, *Quantised Singularities in the Electromagnetic Field*, Proc. Roy. Soc. **133** (1931) 60.

[22] S. Dodelson, *Modern Cosmology*, Academic Press 2003.

[23] P. A. M. Dirac, *Lectures on Quantum Mechanics*, Dover 2001.

[24] M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems*, Princeton 1992.

[25] R. Jackiw, *Quantum meaning of classical field theory*, Rev. Mod. Phys. **49** (1977) 681
R. Jackiw, *Introduction to the Yang-Mills quantum theory*, Rev. Mod. Phys. **52** (1980) 661.

[26] P. Ramond, *Journeys Beyond the Standard Model*, Perseus Books 1999.
R. .N . Mohapatra, *Unification and Supersymmetry. The Frontiers of Quark-Lepton Physics*, Springer 2003.

[27] C. P. Burguess, *Goldstone and pseudogoldstone bosons in nuclear, particle and condensed matter physics*, Phys. Rept. **330** (2000) 193 (`hep-th/9808176`).

[28] L. Álvarez-Gaumé, *An introduction to anomalies*, in: "Fundamental problems of gauge field theory", eds. G. Velo and A. S. Wightman, Plenum Press 1986.
R. A. Bertlmann, *Anomalies in Quantum Field Theory*, Oxford 1996.
K. Fujikawa and H. Suzuki, *Path Integrals and Quantum Anomalies*, Oxford 2004.
J. A. Harvey, *TASI lectures on anomalies*, `hep-th/0509097`.
L. Álvarez-Gaumé and M. A. Vázquez-Mozo, *Introduction to Anomalies*, Springer (to appear).

[29] R. Jackiw, *Topological investigations of quantized gauge theories*, in: "Current Algebra and Anomalies", eds. S. B. Treiman, R. Jackiw, B. Zumino and E. Witten, Princeton 1985.

[30] S. Adler, *Axial-Vector Vertex in Spinor Electrodynamics*, Phys. Rev. **177** (1969) 2426.
J. S. Bell and R. Jackiw, *A PCAC puzzle:* $\pi^0 \to 2\gamma$ *in the sigma model*, Nuovo Cimento **A60** (1969) 47.

[31] J. Steinberger, *On the Use of Substraction Fiels and the Lifetimes of Some Types of Meson Decay*, Phys. Rev. **76** (1949) 1180.

[32] F. J. Ynduráin, *The Theory of Quark and Gluon Interactions*, Springer 1999.

[33] G. 't Hooft, *How the instantons solve the U(1) problem*, Phys. Rept. **142** (1986) 357.

[34] D. G. Sutherland, *Current Algebra and Some Nonstrong Mesonic Decays*, Nucl. Phys. **B2** (1967) 433.
M. J. G. Veltman, *Theoretical aspects of high-energy neutrino interactions*, Proc. R. Soc. **A301** (1967) 107.

[35] S. L. Adler and W. A. Bardeen, *Absence of higher order corrections in the anomalous axial vector divergence equation*, Phys. Rev. **182** (1969) 1517.

[36] E. Witten, *An SU(2) anomaly*, Phys. Lett. **B117** (1982) 324.

[37] S. Eidelman et al. *Review of Particle Phhysics*, Phys. Lett. **B592** (2004) 1 (`http://pdg.lbl.gov`).

[38] D. J. Gross and F. Wilczek, *Ultraviolet behavior of nonabelian gauge theories*, Phys. Rev. Lett. **30** (1973) 1343.

[39] H. D. Politzer, *Reliable perturbative results for strong interations?*, Phys. Rev. Lett. **30** (1973) 1346.

[40] G. 't Hooft, remarks at the *Colloquium on Renormalization of Yang-Mills fields and applications to particle physics*, Marseille 1972.

[41] I. B. Khriplovich, *Green's functions in theories with a non-abelian gauge group*, Yad. Fiz. **10** (1969) 409 [Sov. J. Nucl. Phys. **10** (1970) 235].
M. V. Terentiev and V. S. Vanyashin, *The vacuum polarization of a charged vector field*, Zh. Eskp. Teor. Fiz. **48** (1965) 565 [Sov. Phys. JETP **21** (1965) 375].

[42] K. G. Wilson, *Renormalization group and critical phenomena 1. Renormalization group and the Kadanoff scaling picture*, Phys. Rev. **B4** (1971) 3174.
K. G. Wilson, *Renormalization group and critical phenomena 2. Phase space cell analysis of critical behavior*, Phys. Rev. **B4** (1971) 3184
K. G. Wilson, *The renormalization group and critical phenomena*, Rev. Mod. Phys. **55** (1983) 583.

[43] L. P. Kadanoff, *Scaling Laws for Ising Models Near* T_c, Physics **2** (1966) 263.

[44] J. Schwinger, *On Gauge Invariance and Vacuum Polarization*, Phys. Rev. **82** (1951) 664.

[45] E. Brezin and C. Itzykson, *Pair Production in Vacuum by an Alternating Field*, Phys. Rev. **D2** (1970) 1191.

[46] S. W. Hawking, *Particle Creation by Black Holes*, Commun. Math. Phys. **43** (1975) 199.

[47] M. K. Parikh and F. Wilczek, *Hawking Radiation as Tunneling*, Phys. Rev. Lett. **85** (2000) 5042 (`hep-th/9907001`)

[48] Yu. A. Golfand and E. P. Likhtman, *Extension of the Algebra of Poincaré group generators and violations of P-invariance*, JETP Lett. **13** (1971) 323.
D. V. Volkov and V. P. Akulov, *Is the Neutrino a Goldstone Particle*, Phys. Lett. **B46** (1973) 109.
J. Wess and B. Zumino, *A Lagrangian Model Invariant under Supergauge Transformations*, Phys. Lett. **B49** (1974) 52.

[49] R. Haag, J. Łopuszański and M. Sohnius, *All possible generators of supersymmetries of the S-matrix*, Nucl. Phys. **B88** (1975) 257.