Notes on elementary particle physics

Matteo Giordano

November 2, 2022

Lecture notes for the course "Elementary Particle Physics", Eötvös Loránd University (ELTE), Budapest, year 2022-2023.

Contents

1 Introduction

Elementary particles are (or more precisely are thought to be) the fundamental, indivisible constituents of all matter. Elementary particles come in different types, characterised by several identifying properties such as mass, spin, electric charge, and others to be discussed later. An important characteristic is that all particles of the same type are identical, and so indistinguishable. The aim of elementary particle physics is to study their properties and their interactions, through which macroscopic objects are ultimately built.

The idea that matter is made of indivisible constituents dates back to ancient Greece, most notably to the philosophical work of Democritus and his master Leucippus (around 6th century b.C.). Although this idea proved to be right in the end, it was obviously not based on experimental results, and came from metaphysical speculation only – elementary particle metaphysics, as it were. The first scientifically sound proposal in this direction came from Dalton and his atomic theory at the beginning of the 19th century. Dalton proposed that chemical elements are composed of basic building blocks, the *atoms* (meaning "indivisible" in Greek), which cannot be further divided into smaller pieces, but can be combined forming the various chemical compounds. Contrary to Democritus, he proposed his theory to explain established empirical facts, and the theory was corroborated by further experiments, which led to its refinement. This is how science works, in a nutshell.

In that day and age, atoms were what we would call elementary particles, from which the whole of the matter is built up. As it turned out, however, the indivisible atoms were in fact divisible, and made up of "more elementary" particles. The concept of elementary particle seems therefore to be dependent on the historical period... and in a certain sense it is. Quantum mechanics tells us that to investigate length scales of order Δx we need momenta of the order of $\Delta p \sim \hbar/\Delta x$. The scale down to which one can reach at a given time in history is thus limited by the available energy. While relating elementarity and time is just a (not so funny) pun, relating it to energy makes more sense: if we are interested in a chemistry problem, then it is appropriate to treat atoms as elementary particles. If we are interested instead in problems at higher energies, then the internal structure of the atoms will play a role, and we will have to use electrons, protons and neutrons as elementary particles, and at yet higher energies the internal structure of protons and neutrons will also show up.

Of course, strictly speaking "elementarity" does not come in degrees: something is either elementary or it is not. At the same time, the depth at which we have studied the structure of matter is limited by our current technological capabilities, and what is considered elementary today might not be so tomorrow. Same goes for what we think is right or wrong.[1] Again, this is science at work.

It is worth to get to the list of what are currently considered *the* elementary particles through a brief historical survey. A nice summary of when and how elementary particles were discovered can be found, e.g., in Griffiths' book [1]. Here I will adopt the point of view of what was considered elementary in a certain period of time.

[1] This does not apply to putting pineapple on pizza: that is and always will be an abomination which deserves a special place in Hell.

1.1 Brief history of particle physics

Electrons The birth of elementary particle physics can be considered the discovery of the *electron* (e) by J. J. Thomson in 1897. Experiments with cathode rays had shown that these were deflected by electric and magnetic fields. This suggested that they were actually streams of light, negatively charged particles (from now on, the absolute value of the electron charge will define the unit of electric charge). Thomson determined their speed using (perpendicular) crossed electric and magnetic fields, tuning them so that the stream was undeflected, and making use of the formula $v/c = E/B$ where c is the speed of light. From the deflection of the particles in the presence of the electric field only, he determined their charge-to-mass ratio; from the heat generated by the stream when hitting the screena at the end of the tube, he estimated their energy, and so their mass, which turned out to be about a thousand times smaller than that of a hydrogen atom.

Protons and neutrons Electrons had to come from somewhere, and where else but from atoms? Moreover, their properties turned out to be independent of the material of which the cathode was made, which meant that they had to be present (and the same) in every atom.[2] Since atoms are electrically neutral, that required the presence in the atom of something positively charged. In the period 1908-1917 E. Rutherford (also making use of the experimental results of H. Geiger and E. Marsden) developed his planetary model of the atom, with the electron orbiting around the positively charged nucleus, which contained almost all the mass of the atom. Different atoms contained different nuclei, and a number of electrons equal to the electric charge of the nucleus. He later understood that each unit of positive charge in a nucleus corresponded to one hydrogen nucleus, which came to be known as the *proton* (p). There was still a mismatch between the masses of nuclei and what one would have expected from their charge: this was settled by the discovery of the *neutron* (n) by J. Chadwick in 1932. The neutron forms part of the nuclei, has a mass slightly larger than that of the proton, but is electrically neutral. Decays of unstable nuclei with emission of an electron and increase of their electric charge by one unit (*β decays*) were then understood in terms of one of their neutrons decaying into a proton and an electron.

Photons At this point in time the elementary particles where only four. Indeed, besides the three discussed above, it had become clear that as far as subatomic physics was concerned, the electromagnetic radiation had to be considered as formed by particles, called *photons* (γ). Only in this way it was possible to explain the photoelectric effect (A. Einstein, 1905) and the scattering of light on particles at rest (A. H. Compton, 1926). Photons are electrically neutral and massless; their energy is proportional to the frequency of the radiation.

In terms of photons, the electromagnetic interaction can be seen as the exchange of a stream of photons between the interacting, electrically charged objects (see Fig. 1), carrying energy and momentum (and other quantum numbers as well) from one to the other. This will be the point of view adopted in this course, and later we will describe it in greater detail.

Muons and pions It still remained to be explained how protons could stay together in nuclei despite the strong electric repulsion – and how neutrons could be bound there, too. This

[2]The idea that the atom was not indivisible was actually not new. We tend to forget that often a new understanding of Nature does not come out of the blue, but develops in an already fertile soil.

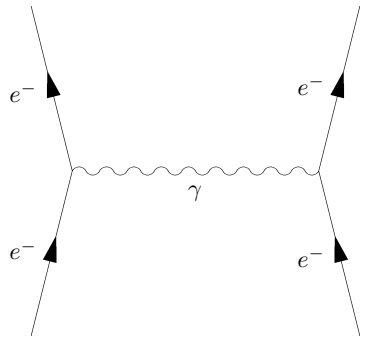

Figure 1: Electromagnetic interactions can be described as exchange of photons between charged particles.

required the existence of a new interaction, called *strong interaction*. This interaction had to be strong to overcome electromagnetic repulsion, but at the same time short-ranged, as it had no effects at macroscopic scales. In analogy with the photon, H. Yukawa proposed in 1934 that the strong interaction was mediated by a new type of massive particle which he called *meson*. The term was coined to distinguish it from the proton and the neutron, which were referred to as *baryons* (from the Greek "heavy"). Together, baryons and mesons form the class of strongly interacting particles, the *hadrons* (from the Greek "strong").[3] From the range of the force, Yukawa estimated the meson mass to be about one sixth of the proton mass. The quest for Yukawa's meson began, and by 1937 a particle with approximately the appropriate mass was found in cosmic rays (C. D. Anderson & S. H. Neddermeyer, 1936)... except it was *not* Yukawa's meson! More detailed studies of cosmic rays in 1946-47 (C. Lattes, G. Occhialini, H. Muirhead, C. F. Powell, 1947) led to conclude that in cosmic rays there were *two* types of particles with masses in the range of interest: one heavier and with a shorter lifetime, which disintegrated almost entirely in the upper atmosphere, and which was the true Yukawa meson; and one lighter and with a longer lifetime, which was what had been initially and incorrectly identified as the meson, but which actually interacted very little with nuclei. The first particle is the *pion* (π), while the second one is the muon (μ), into which the pion mainly decays.[4] While the pion was the missing piece in the strong interaction puzzle, the muon was not expected (and I. I. Rabi famously asked: "Who ordered that?"): it appeared as a sort of heavier electron, and so was grouped with it in the family of *leptons* (from the Greek "small").

Antiparticles In the meantime, it had become clear that each particle had a "twin", called *antiparticle*, with the same mass but opposite electric charge. In fact, after having been predicted theoretically by P. A. M. Dirac, in 1931 the antielectron, or *positron*, had been observed experimentally by C. D. Anderson. The number of elementary particles had grown suddenly by a factor of approximately 2 (some particles, like the photon, are their own antiparticles), even before observations, which came later: in 1955 and 1956 the antiproton and the antineutron[5]

[3] Other exotic objects interacting strongly exists as well. We will ignore them for the time being.

[4] More precisely, pions come in three types, with electric charges ± 1 and 0, and it is the charged ones that decay into muons and antimuons (see below).

[5] For a particle to coincide with its own antiparticle, electric neutrality is a necessary but not sufficient condition: the neutron is different from the antineutron. More on this later.

were also observed, by E. Segrè and O. Chamberlain, and by B. Cork, respectively.

Neutrinos Not all problems had been solved: it remained to be explained why the energy spectrum of the electron in nuclear β decays was extended and not point-like. Energy conservation requires that in this case at least one more particle has to be produced, which in the case at hand has to be neutral and (at least almost) massless. This led to the *neutrino* (ν) hypothesis proposed by W. Pauli in 1930, and incorporated by E. Fermi in 1933 in the first theory of β decay. Direct observation of neutrinos proved to be extremely difficult, since these particles interact very weakly, and it had to wait until the mid-fifties (F. Reines & C. L. Cowan, 1956). Several neutrino experiments were successfully conducted in the '50s and early '60s, and actually revealed that neutrinos and antineutrinos were not the same particle, and, on top of that, that there were two types of neutrinos, one corresponding to the electron (ν_e) and one corresponding to the muon (ν_μ). Neutrinos were included in the lepton family. At this point the interaction responsible for β decays and for processes involving neutrinos, the *weak interaction*, was still poorly understood; strong interactions were not faring much better. We will return on this later.

Strange particles: the quark model and parity violations Everything seemed now to be in place, except possibly for the apparent uselessness of the muon. But the story was far from over. In fact, in 1947 the existence of a new charged particle with mass between that of the pion and that of the proton was confirmed (G. D. Rochester & C. C. Butler): this was the kaon (K), the first of what will become known as "strange" particles, created quickly by strong interactions but decaying slowly via weak interactions. Being affected by the strong interactions, the kaon was included in the hadron family, more precisely in the meson subfamily. More hadrons (both "strange" and not strange) were observed in the following years, leading to a whole new "zoo" of particles. This apparent chaos could finally find some ordering principle when G. Zweig and M. Gell-Mann (independently) proposed the *quark model* in 1964. In this model, mesons and baryons are not elementary, but instead bound states of quarks and antiquarks, a new type of (more...) elementary particles. More precisely, mesons are formed by a quark and an antiquark, and baryons by three quarks (and antibaryons by three antiquarks). Quarks were proposed to exist in three different types: *up* (u), *down* (d), and *strange* (s). This could neatly accomodate all the plethora of strongly interacting particles that had been discovered, and actually correctly predict the existence of a new one, the Omega baryon, observed experimentally in 1964 (V. A. Barnes *et al.*). The story of the success of the quark model is, however, a quite intricate one, and quite surprisingly what convinced most of the community of its validity was the discovery of a fourth type of quark, the *charm* (c), in 1974 (B. Richter *et al.*; S. Ting *et al.*).

Kaons were also responsible for one of the most dramatic discoveries of the 20th century, namely the fact that parity is not a symmetry of nature (Lee and Yang, 1956). A parity transformation is the inversion of all the spatial coordinates, and was for a long time considered a symmetry of nature. A violation of parity means, in a nutshell, that the physics that we see in the mirror is not equivalent to the one that we see here. Experimental signatures of this where sought and found right away (Wu *et al.*, 1957; Garwin *et al.*, 1957).

More leptons and quarks From here on, a third charged lepton, the tau (τ) was discovered in 1975 by M. Perl and collaborators; the corresponding neutrino was immediately theorised,

but observed only in 2000 by the DONUT experiment at Fermilab. A fifth quark was observed in 1977 by L. Lederman and collaborators, the *bottom* or *beauty* (b), which by symmetry reasons led to theorise a sixth one, the *top* or *truth* (t), finally observed in 1995 by the CDF and DØ experiments at Fermilab.

Interactions: the Standard Model The discovery of parity violations discussed above led to a full reconsideration and a better theoretical understanding of weak interactions, with the development of a model involving *intermediate vector bosons* as mediators of the weak force (Glashow, 1960). These are similar to the photon mediating electromagnetic interactions, but massive instead of massless.[6] The resulting model unified the treatment of electromagnetic and weak interactions, taking also into account the intrinsically parity-violating nature of weak interactions. This was later coupled with the recently discovered Higgs mechanism (Higgs, 1964; Brout and Englert, 1964; Guralnik, Higgs, and Kibble, 1964) to yield the currently accepted description of electroweak interactions (Weinberg, 1967; Salam, 1967). The experimental vindication of the Glashow-Salam-Weinberg model required fifty years of intense experimental activity. The intermediate vector bosons, namely the W (charged) and Z (neutral) bosons, were observed in 1983 by the UA1 and UA2 experiments at CERN. The Higgs mechanism requires the existence of a scalar particle known as *Higgs boson* (H), observed in 2012 by the ATLAS and CMS experiments at CERN, whose role will be discussed later.

A few years after the development of the Glashow-Salam-Weinberg model, a successful microscopic model was proposed to describe strong interactions in terms of the exchange of suitable mediators. In fact, the success of the quark model had made clear that the pion could not be considered the fundamental quantum of the strong interactions. This role was taken over by the so-called *gluons* (g), that mediate the interactions between quarks, and also self-interact. Quarks and gluons turned out to carry a further conserved quantum number called *colour*, and for this reason the theory describing their interactions is known as Quantum Chromodynamics (QCD) (Fritzsch, Gell-Mann, and Leutwyler, 1973). However, according to QCD, quarks and gluons are permanently bound together inside hadrons, a property known as *confinement*, which explains why they have not been observed in isolation. Moreover, while gluons are the fundamental mediators of the interaction, due to confinement it is effectively the pions (and other, heavier hadrons as well) that are exchanged to mediate the interaction, thus determining the long-distance properties of the theory. Together, the Glashow-Salam-Weinberg model and QCD form the *Standard Model* of particle physics.

The last of the fundamental interactions is gravity, for which a fully developed quantum-mechanical theoretical formulation does not exist yet. The hypothetical quantum of gravitational interactions, the *graviton* (G), has also not been experimentally observed yet.

1.2 The elementary particles

Let us now summarise what are the known elementary particles (at least, as of now). As it should be clear from the brief history told above, they are organised into two big groups: matter particles and interaction particles. They are listed in Tables 1 and 2. The known nonzero masses of elementary particles span five orders of magnitude, from the 0.5 MeV/c^2 of the electron to the 170 GeV/c^2 of the top quark. This range is expected to widen as soon as neutrino masses will be measured.

[6] The idea of massive mediators of the weak interactions was already proposed by Yukawa.

leptons

particle	charge	mass	spin
e^-	-1	$0.5109989461(31)$ MeV	$\frac{1}{2}$
μ^-	-1	$105.6583745(24)$ MeV	$\frac{1}{2}$
τ^-	-1	$1.77686(12)$ GeV	$\frac{1}{2}$
ν_e	0	< 1.1 eV	$\frac{1}{2}$
ν_μ	0	< 0.19 MeV	$\frac{1}{2}$
ν_τ	0	< 18.2 MeV	$\frac{1}{2}$

quarks

particle	charge	mass(*)	spin
u	$\frac{2}{3}$	$2.16^{+0.49}_{-0.26}$ MeV	$\frac{1}{2}$
d	$-\frac{1}{3}$	$4.67^{+0.48}_{-0.17}$ MeV	$\frac{1}{2}$
s	$-\frac{1}{3}$	93^{+11}_{-5} MeV	$\frac{1}{2}$
c	$\frac{2}{3}$	$1.27^{+0.02}_{-0.02}$ GeV	$\frac{1}{2}$
b	$-\frac{1}{3}$	$4.18^{+0.03}_{-0.02}$ GeV	$\frac{1}{2}$
t	$\frac{2}{3}$	$172.76^{+0.30}_{-0.30}$ GeV	$\frac{1}{2}$

Table 1: Matter particles. Mass values taken from Ref. [2]. Quark masses are the current quark masses in the $\overline{\text{MS}}$ scheme at $\mu \approx 2\,\text{GeV}$ for light quarks, and at the quark mass for the heavy quarks.[7]

1.2.1 Matter particles

Matter particles have a common feature: they are all spin-$\frac{1}{2}$ fermions.[8] They are further subdivided into two groups, *leptons* and *quarks*, each of which is further divided into three *families* (or *generations*).

The leptons include the *electron* (e^-), the *muon* (μ^-) and the tauon or *tau* (τ^-), which are electrically charged, and the corresponding electronic, muonic and tauonic *neutrinos*, which are electrically neutral. To each particle corresponds an antiparticle: the anti-electron or positron, the anti-muon and the anti-tau, and the three anti-neutrinos, which we will collectively call sometimes antileptons. Particle and antiparticle have the same mass and spin, but are distinguished by having opposite electric charge (for the charged ones) or opposite helicities, i.e., spin component in the direction of motion (for the neutrinos). Each charged lepton and its corresponding neutrino form a family, i.e., (e^-, ν_e), (μ^-, ν_μ), (τ^-, ν_τ).

The quarks come in different kinds called *flavours*, named *up*, *down*, *strange*, *charm*, *bottom*

[7] If you do not know what this means do not worry, it is just to make clear that quark masses are a much more delicate issue than it might seem.

[8] For those who forgot. Fermions and bosons are characterised by their symmetry properties under exchange of two of them. The state of a system of identical fermions is antisymmetric under the exchange of any two of them, while the state of a system of identical bosons is symmetric. This in particular implies that no two identical fermions can be in the same state (Pauli's exclusion principle), while there is no such limitation for bosons. Fermions are particles with half-integer spin, while bosons have integer spin. The spin of a particle determines how its states transform under rotations, or more precisely under which representation of the rotation group its states transform. The connection between spin and statistics is a theorem in quantum field theory.

particle	charge	mass	spin
γ	0	0	1
$W^{\pm}$	± 1	80.4 GeV	1
Z	0	91.2 GeV	1
g	0	0	1
H	0	125 GeV	0

Table 2: Interaction particles.

(or *beauty*) and *top* (or *truth*). Quarks are all electrically charged, and paired in families as follows: (u, d), (c, s), (t, b). As with the leptons, with each quark is associated an antiquark.

All matter particles interact through the weak interactions, while only quarks are affected by the strong ones. Quarks and charged leptons interact also electromagnetically, while neutrinos do not.

1.2.2 Interaction particles

Interaction particles are all bosons. They are divided into classes according to what interaction they mediate:

- the photon γ has spin 1, is massless and electrically neutral (i.e., it does not self-interact), and mediates the electromagnetic interactions;
- the intermediate vector bosons $W^{\pm}$ and Z have spin 1, are massive, and mediate the weak interactions; the W's are electrically charged, while the Z is neutral;
- the gluons g have spin 1, are massless, and mediate the strong interactions; they are electrically neutral.

It must be noted that the intermediate vector bosons do interact with each other and with themselves, but not with the gluons; and that the gluons interact with themselves but not with the intermediate vector bosons.

To these particles one has to add the Higgs boson H, which is a massive, electrically neutral particle of spin 0. This interacts with all other elementary particles, except for the photon and the gluons. This interaction provides mass to the elementary particles (and so photon and gluons are massless): we will give a brief explanation later.

The "periodic table" of all the elementary particles is shown in Fig. 2. This does not include antiparticles, that can be easily reconstructed.

1.3 Interactions as particle exchange

As already mentioned above, the point of view adopted in this course is to describe the interaction between particles in terms of the exchange of one or more mediators. In general, particles interacting means that they exchange energy, momentum, angular momentum, and so on. This exchange of quantum numbers is pictured as the exchange of a mediator particle, carrying precisely the right amount of the various quantum numbers from one particle to another. The process of emission or absorption of such a mediator by one of the interacting particles is called an *interaction vertex*. These will be discussed in detail in Section 3.3.

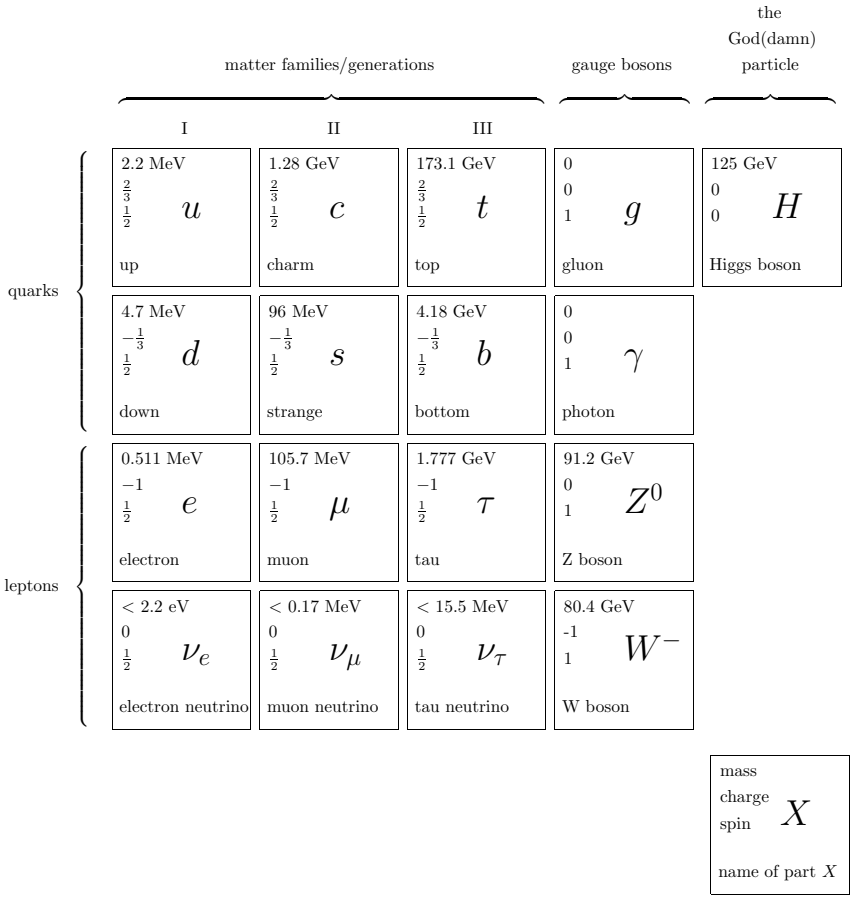

Figure 2: "Periodic table" of elementary particles.

While we distinguished matter and interaction particles above, this does not mean that the latter cannot be observed as well: photons surely can be observed, and to some extent so can the W and Z. Similarly, being the particles associated with the various interactions does not prevent them from interacting themselves. We have already mentioned how matter particles interact by exchanging interaction particles, but the nature of the vertices is such that they can swap roles, with interaction particles exchanging matter particles between them, and in certain cases also other interaction particles of the same or of a different type.

It is easy to show that a process in which a particle a emits a particle b, i.e., $a \to a' + b$ (where a' indicates that the quantum numbers of a have changed), cannot simultaneously conserve energy and momentum.[9] This shows that describing interactions as particle exchange will lead to serious contradictions from a classical perspective. We will see this in more detail later, when we will also clarify what the whole particle-exchange business actually means. In the meantime, though, we can invoke quantum mechanics to come to the rescue through Heisenberg's uncertainty principle. In its hand-waving form, this states that violations of energy conservation are allowed if they take place on time scales satifying $\Delta E \Delta t \geq \frac{\hbar}{2}$. The exchange of a particle with non-trivial quantum numbers is then allowed if it takes place on such a short time scale.

[9]The only exception is the emission of a massless neutral particle of vanishing energy and momentum, which therefore does not carry any quantum numbers and so mediates no interaction.

Range of interactions This argument also allows us to obtain a qualitative estimate of the range of the various interactions, i.e., "how far they are felt" in space. From the uncertainty relation we find that the exchange of a particle of mass M requires to violate energy conservation for a time of order $\Delta t \sim \hbar/\Delta E \sim \hbar/(Mc^2)$, over which the particle can travel a distance not larger than $\Delta x = c\Delta t$ (if it moved at the speed of light). All in all then $\Delta x \sim \hbar/(Mc)$, which is the Compton length of the mediator. This argument is clearly a back-of-an-envelope calculation for several reasons (see Ref. [1], ch. 1, problem 1.2), and we will make it more quantitative later on, but nevertheless it already gives us the correct answer: the range of interaction is the inverse of the mass of the lightest mediator (in appropriate units):

$$\text{range} = \frac{\hbar}{Mc} = \frac{\hbar c}{Mc^2} = \frac{197\ \text{MeV}\cdot\text{fm}}{M[\text{MeV}/c^2]\text{MeV}} = \frac{197}{M[\text{MeV}/c^2]}\,\text{fm}\,. \tag{1.1}$$

Plugging in the W mass, $M_W = 80\,\text{GeV}/c^2$, we find

$$\text{range}_{\text{weak}} = \frac{197}{8\cdot 10^4}\,\text{fm} = 2.5\cdot 10^{-3}\,\text{fm}\,. \tag{1.2}$$

As we have already remarked above, quarks and gluons are confined within hadrons. This entails that only certain combinations of quarks and gluons can be effectively exchanged in a strong process, corresponding to the various hadrons.[10] The lightest mediator, corresponding to the longest interaction range, is the neutral pion with mass $M_{\pi^0} = 135\,\text{MeV}/c^2$, and so the range of strong interactions is

$$\text{range}_{\text{strong}} = \frac{197}{135}\,\text{fm} = 1.5\,\text{fm}\,, \tag{1.3}$$

which is of the sam order as the typical size of nuclei. If we knew instead the interaction range, then we could obtain an estimate of the mass of the lightest mediator, and so of the typical energy scale of the strong and weak interactions, respectively $\mathcal{O}(100\ \text{MeV})$ and $\mathcal{O}(100\ \text{GeV})$. Finally, electromagnetic interactions are mediated by massless photons, which entails an infinite range: indeed, the classical description of the electromagnetic interaction between static charges is provided by Coulomb's potential $V_{\text{Coulomb}} = \frac{e}{4\pi r}$ (in Heaviside-Lorentz units), which contains no length scale.

Strength of interactions In the particle-exchange picture of interactions, their strength is determined by how likely the emission/absorption processes are at the interaction vertices. This is encoded in the so-called *coupling constants* associated with the various vertices: a larger coupling means a higher chance for emission/absorption to take place, and so a stronger interaction between the particles involved in the process.

The type of processes usually considered in particle physics are particle scattering and particle decay (see Fig. 3). In the first case, two particles are thrown against each other: if they happen to pass sufficiently close by, they have a chance to feel each other and deviate from their course, or to be partially or entirely replaced by a different set of particles. The first case is called an *elastic scattering* process, while the second one is called an *inelastic scattering* process. How likely they are to happen depends on how likely the basic emission/absorption processes are.

[10] The idea of hadrons exchanging hadrons was pushed ot its limit in the context of the so-called "bootstrap model" of G. Chew, in which no hadron was considered more elementary or composite than any other ("nuclear democracy"). This approach was abandoned after the discovery of QCD.

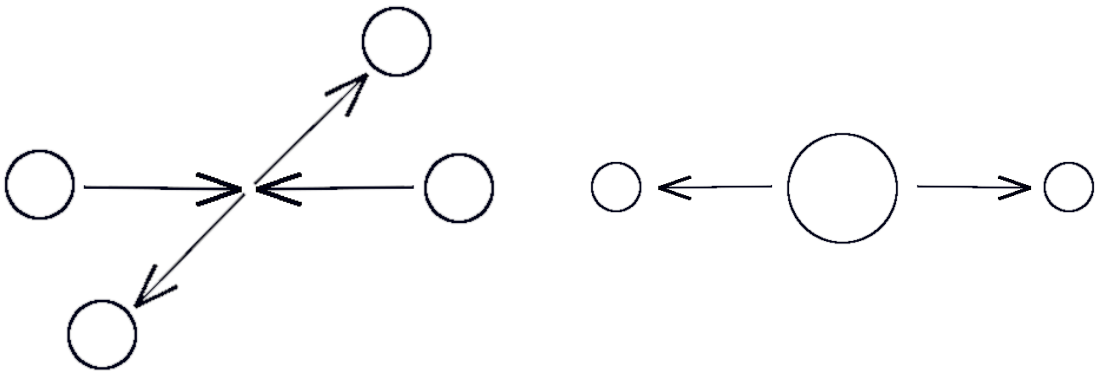

Figure 3: Schematic depiction of a (two-body elastic) scattering process (left) and of a (two-body) decay process (right).

This identifies a sort of surface around the particles, which is more or less "penetrable" without effects, depending on the strength of the interaction. This leads to a typical effective area scale associated with scattering processes, the *cross section*, which takes into account both the geometric component determined by the interaction range, and the strength of the interaction. In most cases it is just one of the various interactions that dominates the scene, so by comparing cross sections from different processes governed by different interactions one can estimate their relative strength. In the case of decay processes, an unstable particle breaks up into other particles. This has an associated time scale, the *lifetime*, telling us for how long the unstable particle is likely to survive before decaying. In turn, the inverse of the lifetime tells us the rate (number of decays per unit time) at which a sample of unstable particle decays. Again, in most cases the decay process is dominated by one of the interactions, and the stronger this is the larger the decay rate (and so the shorter the lifetime) will be.[11] Comparing lifetimes allows us again to estimate the relative strength of the various interactions.

1.4 Natural units

You may have noticed that the masses of the elementary particles in Tables 1 and 2 are expressed in energy units. This is possible because of the existence of a fundamental constant, the speed of light c, that allows one to translate masses into energies. Recalling the Einstein relation $E = mc^2$, we see that a correct mass unit in the usual system is $[m] = \mathrm{eV}/c^2$. Nothing would prevent us though from reporting mc^2 instead of m, if we so wished. Similarly, the existence of the Planck constant $\hbar$ allows one to express times in terms of inverse energies: recalling here the relation between the energy of a photon and its frequency, $E = \hbar\nu$, we see that in the usual system $[t] = \hbar/\mathrm{eV}$ is a correct time unit. Using together c and $\hbar$ we can express lengths in units of $[l] = \hbar c/\mathrm{eV}$. Combining this with the relation between energy and mass we can write a length in units of $\hbar c/(c^2 \cdot \mathrm{eV}/c^2) = (\hbar/c)/(\mathrm{eV}/c^2) = (\hbar/c)/[m]$. Finally, from the Coulomb

[11] One way to see it is to imagine to reverse the process, throwing the decay products against each other and counting how many unstable particles are produced. Up to factors, the original decay rate is turned now into the production rate, and a stronger interaction leads to a higher production rate – so a larger decay rate and a shorter lifetime in the original problem. Notice also that it is not only the strength of the interaction that matters enters, but other features as well, such as the mass difference between the parent and daughter particles.

	quark content	spin	charge	mass
meson				
π^+	$u\bar{d}$	0	$+1$	140 MeV
π^-	$d\bar{u}$	0	-1	140 MeV
π^0	$u\bar{u}, d\bar{d}$	0	0	135 MeV
ρ^+	$u\bar{d}$	1	$+1$	775 MeV
ρ^-	$d\bar{u}$	1	-1	775 MeV
ρ^0	$u\bar{u}, d\bar{d}$	1	0	775 MeV
K^+	$u\bar{s}$	0	$+1$	494 MeV
K^-	$s\bar{u}$	0	-1	494 MeV
K^0	$d\bar{s}$	0	0	498 MeV
$\bar{K}^0$	$s\bar{d}$	0	0	498 MeV
J/ψ	$c\bar{c}$	1	0	3.1 GeV
baryon				
p	uud	$\frac{1}{2}$	$+1$	0.938 GeV
n	udd	$\frac{1}{2}$	$+1$	0.940 GeV
Δ^+	uud	$\frac{3}{2}$	$+1$	1.232 GeV
Δ^0	udd	$\frac{3}{2}$	$+1$	1.232 GeV
Λ	uds	$\frac{1}{2}$	0	1.1 GeV

Table 3: Hadrons. Masses are very different from the quark masses (for light mesons), the difference being accounted for by the interactions.

potential energy (in Heaviside-Lorentz units) $U = e^2/(4\pi r)$ we find $[e]^2 = [E][l] = [\hbar c]$. Also in these cases we could report times, lengths, and electric charges after multiplying them with the appropriate powers of c and $\hbar$.

In particle physics it is convenient to choose our system of units such that $\hbar = c = 1$: this is the so called system of *natural units*. In this system a length has dimensions of an inverse mass; mass and energy have the same dimensions, and so do time and length; finally, electric charge is dimensionless. For example, the length scale associated with a particle is the Compton length $\lambda_C = \frac{\hbar}{mc}$, that we saw above when discussing the range of interactions; in natural units this is just $\lambda_C = \frac{1}{m}$. The *fine structure constant* $\alpha = e^2/(4\pi\hbar c)$ is dimensionless, and in natural units it reads simply $e^2/(4\pi)$. Notice that the two fundamental constants c and $\hbar$ originate respectively from special relativity and quantum mechanics, that both play a role in the description of high-energy but microscopic processes such as those studied in particle physics.

The typical energy unit used in particle physics is the electronvolt (eV). An electronvolt eV $= 1.6 \cdot 10^{-19}$ J is the energy acquired by an electron after travelling through an electric potential difference of one volt: this turns out to be a very convenient unit in accelerator experiments. Practically useful units of energy and length are the megaelectronvolt, 1 MeV $= 10^6$ eV and the gigaelectronvolt, 1 GeV $= 10^9$ eV; and the fermi, 1 fm $= 10^{-15}$ m. In natural units, conversion between fm and eV^{-1} is most easily done by exploiting the relation $\hbar c \simeq 197$ MeV $\cdot$ fm: in natural units the left-hand side is one, so 1 fm $\approx (1/5)\,\text{GeV}^{-1}$.

1.5 Building up matter

From the elementary particles in Tables 1 and 2 one can ultimately build up all the matter surrounding us, all the way from the electron and the proton to that particularly dirty yellow bathroom towel that your cousin never changes (yes Gary, you know which one). This happens in several stages, characterised by different length and energy scales.

The first stage is the construction of hadrons from quarks and gluons (see Table 3). As we have already mentioned, it is a fact of life that free quarks and gluons are not observed in Nature, but always come along bounded into mesons and baryons due to confinement. To a first approximation, mesons are bound states of a quark and an antiquark, while baryons are made up of three quarks (and antibaryons of three antiquarks).[12]

The lightest meson is the pion, that comes in three versions: π^0, π^+ and π^-. The pions are built from the lightest quarks and antiquarks, the u, d, $\bar{u}$ and $\bar{d}$, combined into states of vanishing total spin and orbital angular momentum, so that they have spin 0. The same quark content can lead to different mesons if quarks are in a different spin and/or orbital angular momentum state: for example the ρ mesons have the same quark content of the pions, but in a combination with total spin 1 and vanishing orbital angular momentum, so that they have spin 1. Of course, other mesons can be built changing the quark content: this is the case of the kaons, which contain a strange quark, or of the J/ψ, which is a $c\bar{c}$ state.

The lightest baryon is the proton, made up of two u and one d quark, followed by the neutron with two d and one u quark. Similarly, the lightest antibaryons are the antiproton, made up of two $\bar{u}$ and one $\bar{d}$ antiquark, and the antineutron made up of two $\bar{d}$ and one $\bar{u}$ antiquark – and so different from the neutron, as anticipated. The neutron is slightly more massive than the proton, and it actually decays into it via β decay. The proton on the other hand is stable (and luckily so), precisely because it is the lightest baryon: as a matter of fact, the *baryon number*, i.e., the number of baryons minus the number of antibaryons, is a conserved quantity, and having no other baryon to decay into, the proton sits quietly at the centre of the hydrogen atom, or together with other protons and neutrons[13] in the nucleus of heavier atoms, guaranteeing the stability of ordinary matter. On the contrary, there is no such a thing as a conserved meson number; in fact, even the pion is not stable and decays (mostly) into a muon and a muonic antineutrino, or an antimuon and a muonic neutrino, for the charged types; and into two photons for the neutral type. As with the mesons, heavier baryons exist with the same quark content but in different internal states, and of course with different quark content. An example of the first case are the Δ^+ and Δ^0, which have the same quark content as the proton and the neutron, respectively, but spin $\frac{3}{2}$. An example of the second case is the Λ, which contains a u, a d and an s quark. We will discuss in detail the taxonomy of mesons and baryons in Section 5.

As mentioned above, protons and neutrons form the nuclei of atoms, where they are bound together by the strong interaction, i.e., by the exchange of gluons and quarks, or effectively by pions and other hadrons. Nuclei are the second stage of the construction of matter. They come in a large variety, characterised by their electric charge (i.e., the number of protons) and mass (i.e., the number of proton and neutrons, up to a proportionality factor and small corrections), and they can be stable or not. The study of nuclei is the subject of nuclear physics.

The next stage is the formation of atoms, out of some nucleus and of electrons bound to

[12] A more refined description should take into account also how many gluons and quark-antiquark pairs are present. We will content ourselves with the simplest picture.

[13] Neutrons can be stable when part of nuclei, although not in all cases.

particle	main decay mode	lifetime
μ^-	$\mu^- \to e^- \nu_\mu \bar{\nu}_e$	$2.2 \cdot 10^{-6}$ s
n	$n \to p e^- \bar{\nu}_e$	$8.8 \cdot 10^{2}$ s
π^+	$\pi^+ \to \mu^+ \nu_\mu$	$2.6 \cdot 10^{-8}$ s
π^0	$\pi^0 \to \gamma\gamma$	$8.4 \cdot 10^{-17}$ s
Δ^0	$\Delta^0 \to p\pi^-, n\pi^0$	$5.6 \cdot 10^{-24}$ s

Table 4: Decays of unstable particles.

it by electromagnetic forces. The electron is the lightest charged particle and is stable (again, luckily for us) thanks to conservation of electric charge. Atoms are electrically neutral, with as many electrons as protons in the nucleus; atoms differing only by the numbers of neutrons in the nucleus are called *isotopes*. Atoms with one or more electrons missing, or in excess, are called *ions*. The study of atoms is the subject of atomic physics. Also all the following stages up to everyday's scales are governed by electromagnetic interactions: these are responsible for binding atoms together in molecules, and molecules together in more and more complicated structures, leading ultimately to apples, cows, humans, or Donald Trump. Larger scales (planetary and above) involve gravity, and are outside of the scope of this course.

1.6 Unstable particles and decays

The proton, the electron, the neutrinos,[14] the photon and their corresponding antiparticles, are the only stable particles.[15] Other hadrons and leptons, as well as the intermediate vector bosons,[16] *decay*, i.e., they "break up" in various ways,

$$X \to X_1 X_2 \dots X_n \,, \tag{1.4}$$

yielding ultimately the stable particles listed above. The typical mean time τ that it takes for a particle to decay is called *lifetime*: this is $\tau = t_{\frac{1}{2}} / \ln 2$ where the *half-life* $t_{\frac{1}{2}}$ is the time it takes for half of a large sample of particles to decay. A list of the main decay modes of a few unstable particles and their lifetimes is given in Table 4.

Notice that the classifications in elementary and composite, stable and unstable, or observable and unobservable are quite independent: particles can be elementary, stable and observable (e.g., the electron), elementary, unstable and observable (e.g., the muon), composite, stable and observable (e.g., the proton), elementary and unobservable (e.g., the up quark in a proton), and so on.

The meaning of lifetime is related to the empirical (approximate, but rather accurate) finding that the decay rate of unstable particles, i.e., the probability of decaying per unit time per particle, is independent of time and of sample size (at least for sufficiently big samples). If dP is the probability for a particle to decay over an infinitesimal time interval dt, one then writes $dP = \Gamma dt$ with time-independent Γ. The quantity Γ is the *decay width* of the particle. Since

[14] For simplicity we will treat the neutrinos as massless, which guarantees their stability, even though we know that they are not. Nevertheless, even in this case there is a linear combination of neutrinos which is the lightest in mass, and which is the lightest lepton: its stability is then guaranteed by lepton number conservation (see below).

[15] We are not including nuclei in the discussion here, since we look at them as composite particles at a higher level. Also neutrons, which can become stable in nuclei, are not included.

[16] Like quarks, gluons do not exist as isolated particles.

each particle decays independently of the others (at least for sufficiently big samples), the size $N(t)$ of the sample as a function of time obeys the equation

$$dN(t) = -N(t)dP = -\Gamma N(t)dt\,, \tag{1.5}$$

and so one finds the exponential decay law

$$N(t) = N(0)e^{-\Gamma t} = N(0)e^{-\frac{t}{\tau}}\,, \tag{1.6}$$

where we have identified the lifetime τ as $\tau = \Gamma^{-1}$. The relation between τ and the half-life $t_{\frac{1}{2}}$ then follows.

Typically, unstable particles decay through different decay modes, yielding different final products. Each of these modes is a *decay channel*, and the i-th channel is characterised by the *partial width* Γ_i, with the total width being given by $\Gamma = \sum_i \Gamma_i$. The ratio Γ_i/Γ is the *branching ratio* (or *fraction*) of the i-th channel, and tells us the relative probability that the decay will take place through channel i.

In general, not all ways of decaying are allowed, due to kinematical and dynamical constraints. The most important kinematical constraint stems from energy-momentum conservation. Denoting with P the four-momentum of the initial particle, of mass M, and with p_i those of the decay products, of masses m_i, we have from $P = \sum_i p_i$ that

$$M^2 = P^2 = P \cdot \sum_i p_i = M \sum_i E_i^{\rm CM} \geq M \sum_i m_i\,, \tag{1.7}$$

where CM denotes the rest frame of the initial particle and $E_i^{\rm CM} = \sqrt{(\vec{p}_i^{\,\rm CM})^2 + m_i^2}$. This means that the sum of the masses of the decay products cannot exceed that of the initial particle.

Other constraints of dynamical nature come about because of how the various interactions work, and lead to further interaction-specific conservation laws. Examples are conservation of electric charge, baryon number, lepton and (approximately) lepton family number (see below).

1.7 How to tell the nature of a process: decays and conservation laws

We have mentioned above that by comparing cross sections or lifetimes of processes governed by different interactions we can estimate their relative strength – but how can we tell that two processes are governed by different interactions, and decide which one is governed by which?

Focussing on decay processes, it is an experimental finding that lifetimes fall in three clearly separated ranges. At the fast end, there are processes with lifetimes in the range $\tau \sim 10^{-23} \div 10^{-20} s$; in the middle there are those with $\tau \sim 10^{-16} s$; and at the slow end there are those with $\tau \sim 10^{-13} \div 10^{3} s$. From our considerations above, this indicates that the interactions causing the processes in the three sets have clearly different and increasing strengths. This already tells us that the interactions governing the three sets of processes must be different. It would be natural to call strong the interactions behind the fast processes, and weak those behind the slow ones: as we will see momentarily, these correspond precisely to what we called strong and weak interactions in our introductory discussion, leaving the electromagnetic interactions as the ones responsible for the intermediate processes.

In fact, there is a second feature that is common to many processes in the same set, namely the presence of "signature particles": the slow processes are often characterised by the presence of

neutrinos or antineutrinos; the intermediate ones often produce photons; and the fast ones often produce pions. This should reassure you that our classification based on interaction strength matches our previous discussion, with the hierarchy strong > electromagnetic > weak.

Given now a process, we can reverse the reasoning and use signature particles and lifetimes to identify the relevant interaction: if we see a photon or a neutrino coming out of a process we can be sure that those processes involve the electromagnetic interaction or the weak interaction, respectively, and if we see pions (without neutrinos) there are good chances that we are dealing with a strong process (although we cannot be entirely sure). In the absence of signature particles, we can still rely on the strong hierarchy in the strength of the various interactions, reflecting into the very different orders of magnitude for the lifetimes of the decays that they govern. On the other hand, we may not know the lifetime of the process; or we may be wondering if a process may or may not take place, and if so what would be the relevant interaction.[17] In this case we have a powerful tool at our disposal, namely conservation laws, that we know discuss. While we keep focussing on decay processes for simplicity, similar considerations can be made for scattering processes as well.

As we have already mentioned, particles generally decay into lighter particles, unless this is forbidden by some conservation law. Similarly, certain inelatic processes are forbidden due to the mismatch of some conserved quantity on the two sides of the reaction. Certain conservation laws are valid for all interactions, namely conservation of energy, momentum, and angular momentum, that imply kinematical constraints on the decay processes. For example, as we proved above, energy conservation implies that for a decay process to be allowed the mass of the products cannot exceed the mass of the initial particle. Other conservation laws are instead of dynamical nature, deriving from the details of the various interactions. In general, conservation laws are associated with the symmetries of the system: for example, energy, momentum, and angular momentum conservation are consequences respectively of invariance under temporal translations, spatial translations, and rotations, which are very general properties of any system of interacting particles.

Among the dynamical conservation laws, some are universally valid: for example, we have already mentioned above that all interactions conserve electric charge, baryon number and lepton number. A process violating any of these laws is immediately forbidden. Other conservation laws are instead valid for certain interactions and not for others, and this allows one to identify what is the interaction responsible for a given process where such conservation laws are violated, by excluding those interactions that do not allow any such violation. We want to stress that although we have mostly been discussing decay processes, of course these and other conservation laws do apply to scattering processes as well.

While we will discuss symmetries in greater detail later on, here we want to give a brief overview of the symmetries of the various interactions. Strong interactions are certainly the "most symmetrical" ones, respecting all the important symmetries exactly and some further symmetries approximately, followed by electromagnetism which again exactly respects all the important symmetries, and by weak interactions that violate a good number of them.

Particle-type conservation laws We begin with particle-type conservation laws, i.e., regarding the conservation of the number of particles minus number of antiparticles of a specific type, or of a specific subset of particle types. Both strong and electromagnetic interactions conserve

[17] If several are available, it is the strongest one that dominates.

each individual particle number, while weak interactions conserve only certain combinations of these numbers, so it is useful to establish some nomenclature.

To each of the different types (flavours) of quarks is associated a flavour quantum number: "upness", $U = u - \bar{u}$, "downness", $D = \bar{d} - d$, "strangeness" $S = \bar{s} - s$, "charm" $C = c - \bar{c}$, "beauty" $B = \bar{b} - b$, and "topness" $T = t - \bar{t}$.[18] Among these numbers, topness is of scarce relevance, as no hadrons containing t or $\bar{t}$ exist (the top quark decays too fast for a bound state to form). Moreover, upness and downness are never used, as they are usually traded for the quark number $\mathcal{Q}$ (total number of quarks minus antiquarks) or equivalently and more commonly for the baryon number $\mathcal{B} = \mathcal{Q}/3$ (recall that a baryon contains three quarks, and an antibaryon three antiquarks); and for the hadronic component of the electric charge $Q_h = \frac{2}{3}(U + C + T) + \frac{1}{3}(D + S + B)$. All flavour quantum numbers are conserved by strong and electromagnetic interactions, and so *a fortiori* are $\mathcal{B}$ and Q_h, while only the total baryon number $\mathcal{B}$ is conserved by weak interactions.

Leptonic flavour numbers are similarly (but more regularly) assigned to the various types of leptons as number of leptons minus number of antileptons; as above, one can trade the flavour number of one of the charged types for the leptonic component of the electric charge, $Q_l = -(e + \mu + \tau)$. All of these are individually conserved by electromagnetic interactions, and trivially by strong interactions as well, but again not conserved by weak interactions. On the other hand, if we count together leptons in the same family by assigning *lepton family number* $L_{e,\mu,\tau} = 1$ to e^-, μ^-, τ^- and ν_e, ν_μ, ν_τ, respectively, and $L_{e,\mu,\tau} = -1$ to e^+, μ^+, τ^+ and $\bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$, respectively, and zero in any other case, then the total lepton family numbers (i.e., the sum of the lepton family number of each particle in the system, and so the total number of leptons minus antileptons of a given family) are conserved also by weak interactions. While a similar "quark family" number could be defined by combining the flavour numbers of quarks in the same family, this would be of very limited usefulness, as it is does not add any information in the strong and electromagnetic case, and it is not conserved in the weak case. Finally, while Q_h and Q_l are not individually conserved by weak interactions, the total electric charge $Q = Q_h + Q_l$ is.

Nowadays it is known that the lepton family numbers are actually not exactly conserved. In fact, this would be the case if neutrinos were exactly massless, but for nonzero masses it is possible for neutrinos to "oscillate", changing their type, e.g., from electronic to tauonic, thus violating family lepton number conservation. Such oscillations have actually been observed, proving that neutrinos have nonzero masses. Nonetheless, in this course we will mostly treat them as massless for simplicity, as this describes accurately most of the important physics. On the other hand, the total *lepton number* $L = L_e + L_\mu + L_\tau$ remains conserved by weak interactions also in the presence of neutrino oscillations.

A summary of the conservation laws related to particle types is given in Table 5. Their origin will become apparent in Section 3.3. An obvious consequence of these conservation laws is that if an interaction conserves a certain particle number, then it cannot be responsible for decays in which this number is violated: for example, strong interactions cannot be responsible for strangeness-changing processes.

[18] Notice the different choices of signs, due to historical reason: when the quark model was formulated, the empirical assignement of the strangeness quantum number to hadrons turned out to correspond to the number of strange antiquarks minus number of strange quarks. The flavour quantum numbers of the other negatively charged quarks are defined in analogy with this.

	interaction		
quantity	strong	electromagnetic	weak
lepton type	✓ (trivially)	✓	✗
lepton family L_ℓ	✓ (trivially)	✓	✓ (if ν_ℓ are massless)
lepton number $L = \sum_\ell L_\ell$	✓ (trivially)	✓	✓
flavour U, D, S, C, B, T (equivalently $U, D \to \mathcal{B}, Q_h$)	✓	✓	✗
quark family (not really useful...)	✓	✓	✗
quark number $\mathcal{Q}$ or baryon number $\mathcal{B} = \mathcal{Q}/3$	✓	✓	✓
electric charge Q	✓	✓	✓

Table 5: Conservation laws (I): particle-type quantum numbers.

Approximate flavour symmetries Strong interactions also conserve approximately a quantity called *isospin*. We will see later that this is a very good approximate symmetry, that originates from the fact that only the quark mass distinguishes the different quark flavours as far as strong interactions are concerned: since the two lightest quarks (u and d) have similar masses, they are basically indistinguishable for strong interactions. One can then "rotate" one flavour into the other without physical effects. This rotational symmetry is mathematically identical to the spacetime one, from which the concept of spin originates. From the physical point of view, however, it is very different and totally unrelated: in fact, these flavour rotations act in the internal space spanned by the up and down component of the quark state, and do not affect spacetime in any way. For this reason, isospin is called an *internal symmetry*. Analogously to the angular-momentum multiplets of degenerate energy levels found in atomic spectra due to rotation symmetry, isospin manifests through the appearance of multiplets of (nearly) degenerate light hadrons (i.e., those made of u, d and s). Borrowing from the theory of spin, these multiplets can be classified by assigning them a total isospin number I, with $2I + 1$ being their degeneracy, and further distinguishing its members by a second number $I_3 = -I, \ldots, +I$, chosen to increase with their electric charge. All other particles (non-light hadrons, leptons, etc.) are assigned $I = I_3 = 0$. To list a few examples:

- the three pions form a triplet to which it is assigned the value $I = 1$, and to each of them one assigns I_3 equal to their charge, $I_3 = Q$;
- the proton and the neutron form a doublet with $I = \frac{1}{2}$ and $I_3 = Q - \frac{1}{2}$;
- the K^+ and K^0 form again a doublet with $I = \frac{1}{2}$ and $I_3 = Q - 1$;
- the Δ resonances Δ^{++}, Δ^+, Δ^0 and Δ^- form a quartet with $I = \frac{3}{2}$ and $I_3 = Q - \frac{1}{2}$.

In general $I_3 = Q - \frac{1}{2}(\mathcal{B} + S)$, so I_3 is automatically conserved since $Q, \mathcal{B}$ and S are. However, also $I^2 = I_1^2 + I_2^2 + I_3^2$ is conserved, and as we will see this has interesting consequences for

		interaction		
quantity		strong	electromagnetic	weak
isospin	I	✓	✗	✗
discrete symmetries	P	✓	✓	✗
	C	✓	✓	✗
	T	✓	✓	✗
	CP	✓	✓	✗
	CPT	✓	✓	✓

Table 6: Conservation laws (II): isospin and discrete symmetries.

scattering processes. Both electromagnetic and weak interactions do *not* conserve this quantity (pion decay modes should convince you of that).

If one relaxes how accurate the approximate symmetry has to be, one can extend the discussion from the two lightest quarks to the three lightest quarks, and so on, possibly up to the full set of flavours. These *flavour symmetries* are less and less accurate, but still provide useful information on strongly interacting matter. In particular, in Section 5.2.5 we will discuss in detail the approximate flavour symmetry involving the three lightest quarks.

Discrete symmetries Finally, there are three important discrete symmetries to mention, namely *parity* P (i.e., spatial inversion), *charge conjugation* C (exchange of particle with antiparticle) and *time reversal* T (i.e., inversion of the direction of time). Strong and electromagnetic interactions conserve all three of them, while weak interactions do not conserve any of them taken individually. Also relevant are the combinations CP and CPT: the first is conserved by weak interactions to some extent but not exactly, while CPT is conserved exactly, so making CP and T equivalent symmetries. Conservation of the combination CPT is expected on very general theoretical grounds, and experimental observations of CPT violations would require a radical revision of our theoretical description of particle physics.[19]

The status of the main symmetries for the various interactions is summarised in Tables 5 and 6. As already pointed out, they can be used to identify by exclusion what interaction is responsible for a process. For example, if we see a process that does not conserve neither isospin nor strangeness we know that it is happening via weak interactions; if it conserves strangeness but not isospin then it happens via electromagnetic interactions; and so on. If it conserves everything, then it is most likely taking place through the strong interactions, but we cannot exclude the others: non-conservation of a quantum number does not imply that it is always not conserved in a process.

2 Scattering processes

If unstable particles decay, how are we even able to know about their existence? For some of them, we might argue that if they have been created when the Earth (or even the Universe

[19] Currently, particle interactions are successfully described in the framework of Quantum Field Theory, where CPT conservation is a general theorem.

itself) appeared, and their lifetime is long enough, then part of them might have survived this long and can still be observed. This is the case, e.g., for uranium nuclei. However, this does not fully answer our question, since we know of particles with lifetimes much shorter than the age of the Universe, and indeed much shorter than the lifespan of a cat on a highway. How can we possibly know about them?

Imagine to look at an unstable particle decay, breaking up into a certain number of decay products. If we reverse the direction of time, what we would see is a set of particles getting closer, colliding, and building up the unstable particle. After all, the laws of physics are (to a good extent) invariant under time reversal, and the process resulting from "projecting the movie backwards" would be a possible physical process. Here is then our answer to our initial question: to see unstable particles we have to create them, and this can be achieved by means of *scattering processes*, in which particles are thrown at each other, and the products of the collision are studied. If an unstable particle is created in the processes, the final products will show its distinctive footprint. Scattering processes are actually much more than just a tool to find out the spectrum of unstable particles, as they reveal important information on how the interactions between particles work. For this reason, scattering experiments are the main type of experiment in particle physics.

Scattering experiments are set up by throwing a bunch of particles against a fixed target or against another bunch of particles, and carefully analysing what comes out of the collision. Such experiments are typically arranged so that only one pair of particles is interacting at a time. It is an experimental fact that particle states can be prepared so that the particles are far enough from each other that they do not interact appreciably, thus travelling essentially undisturbed on straight-line trajectories. This is justified a posteriori by the fact that interactions are typically short-ranged.[20] These states are typically used as the *initial states* of scattering experiments. It is another experimental fact that the state of the system, after a sufficiently long time has elapsed after the collision, looks again like a state of freely-evolving particles. What is a sufficiently long time in a scattering experiment depends on the type of interaction, but it is safe to say that no matter how long it is, it is a very short time on human scales: an estimate of $10^{-10}s$ for an upper bound is reported in Taylor's book [3]. For all practical purposes we can then imagine that the system is prepared in its initial state in the distant past (formally $t = -\infty$), when particles are far away from each other, and that observations are made in the distant future (formally $t = +\infty$), when particles are again far away from each other after having interacted. Observations are made by means of detectors that measure energy, momentum, electric charge, etc., of the final particles, and which therefore (in the language of Quantum Mechanics) project the state of the system on some particle state with definite particle content and particle momenta.

2.1 Scattering experiments

The two main types of scattering experiment are *fixed-target* and *collider* experiments.

Fixed target In a fixed-target experiment, a bunch of particles is accelerated and focussed into a *beam*, which is then directed against a target, for example a thin foil of metal, at rest in the laboratory (see Fig. 4). What comes out of the collision is carefully analysed by means of detectors placed behind (or around) the target. The beam is characterised by the number N_b of

[20] Even the long-ranged electromagnetic interactions are in most of the cases effectively short-ranged in practice, due to screening effects.

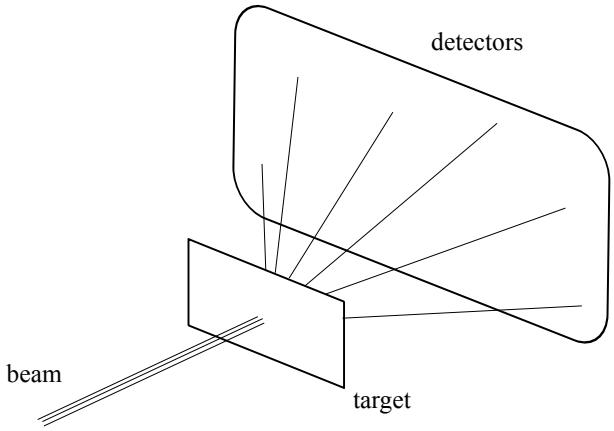

Figure 4: Setup of a fixed-target scattering experiment.

particles that it contains, by their velocity v, and by its cross-sectional area A_b. The target is characterised by its density and thickness, from which one obtains the number N_t of particles in that part of the target on which the beam impinges. Experiments are typically designed so that the target is much wider than the beam cross-section.

Since interactions are typically of short-range nature, a particle in the beam will interact with a particle in the target only if it gets close enough to it to "feel" it. This effectively defines a region around the target particle ("scatterer") in the plane orthogonal to the trajectory of the beam particle ("projectile") in which this trajectory has to pass, if a scattering event is to take place. The strength (and other details) of the interaction further affects how likely this is to happen, making this region more or less penetrable. The effective area of this region, which combines its geometric size with its "opacity", measures how likely it is for a scattering event to take place, once that the technical details mentioned above about the beam and the target are known. In fact, for sufficiently dilute beams and targets, interactions will involve at most one projectile and one scatterer, and so scattering events will be independent: their number will be proportional to the number of particles in the beam and in the target. In particular, if the target is thin enough, a projectile will not undergo multiple scattering processes: from its point of view the target then looks like a two-dimensional surface with separated "active" areas corresponding to the scatterers, and so the target can effectively be imagined as being collapsed on a single two-dimensional sheet (see Fig. 5). For each projectile, that we are assuming can be found with uniform probability anywhere on the beam cross-section, the probability of being scattered is then equal to the ratio of the active area under the beam ($\propto N_t$) divided by the area of the beam cross-section (A_b).

All in all, the number of scattering events N_{events} will be proportional to N_t and to N_b (once the whole beam has passed though the target), and inversely proportional to A_b. The proportionality factor is the *total cross section*, σ, of the scattering process, as it depends uniquely on the type of particles involved (besides of course their energy and spin components). The cross section has dimensions of an area, not surprisingly since it measures the effective area of the "active" region around each scatterer in the target, as discussed above. In formulas,

$$N_{\text{events}} = N_t N_b \frac{\sigma}{A_b}\,, \tag{2.1}$$

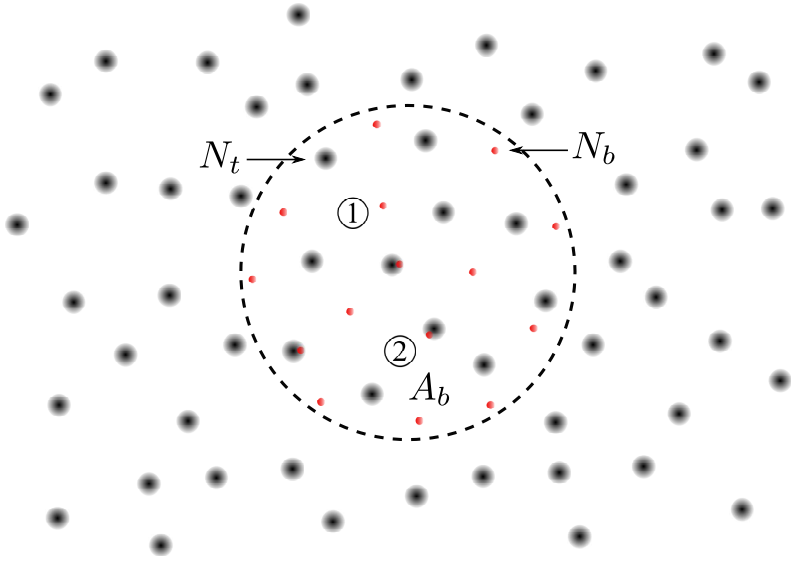

Figure 5: The target as seen from the beam. Target particles are the black discs, with area corresponding to the cross section σ of the process, beam particles are the red dots. The dashed line indicates the cross-section of the beam, of area A_b, N_b is the number of particles in the beam, and N_t is the number of particles in the target "seen" by the beam. The beam particle 1 does not undergo scattering, while particle 2 does.

which when turned around gives the operative definition of the cross section,

$$\sigma = \frac{N_{\text{events}}}{N_t \frac{N_b}{A_b}} \,. \tag{2.2}$$

For practical purposes it is convenient to elaborate on this formula. The number of particles in the beam can be obviously expressed as the number of the particles that cross the target per unit time, times the time it takes for the whole beam to cross the target. Assuming for simplicity a uniform longitudinal distribution of particles in the beam and a constant velocity v, the number of beam particles ΔN_b crossing the target in the small time interval $(t, t+\Delta t)$ are those that are at most a small distance $\Delta x = v\Delta t$ away from the target. If ρ_b is the volume density of particles in the beam, then $\Delta N_b = \rho_b A_b \Delta x = \rho_b A_b v \Delta t$. The number of events ΔN_{events} happening in Δt will be given by Eq. (2.1) with N_b replaced by ΔN_b, and so in Eq. (2.2) we can replace $\frac{N_{\text{events}}}{N_b/A_b}$ with $\frac{\Delta N_{\text{events}}}{\Delta N_b/A_b} = \frac{\Delta N_{\text{events}}}{\Delta t \rho_b v}$. The quantity $\Phi \equiv \rho_b v$ measures the number of particles of the beam crossing the target per unit time and unit area (on the plane perpendicular to the beam velocity), and is called the *flux* of the beam. We can then write

$$\sigma = \frac{1}{N_t \Phi} \frac{\Delta N_{\text{events}}}{\Delta t} \,. \tag{2.3}$$

i.e., the cross section is the number of scattering events per unit time (or *event rate*) per unit target and unit flux.

As we already pointed out, the larger σ, generally stronger the interaction: a larger cross section corresponds to a larger probability of interaction, and so to a stronger force. This is

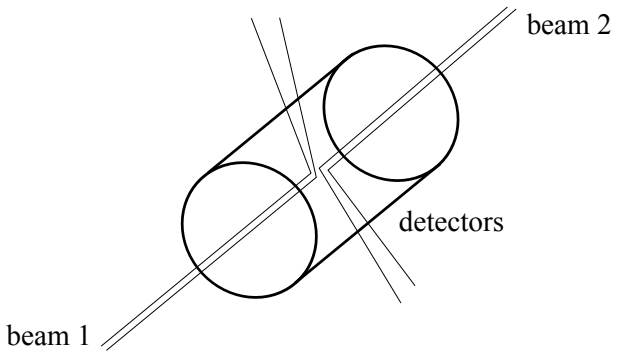

Figure 6: Collider experiment.

analogous to what happens with the decay rate Γ, and one can similarly use σ to estimate the strength of the various interactions.

Instead of counting all the scattering events, one can classify them according to the type and number of final particles, their momenta and polarisation, and so on, and count how many events with prescribed features take place. Discrete variables, like number and type of particles involved and their polarisations, essentially label different scattering processes: let us denote them collectively with the symbol α. Let ξ instead denote collectively the continuous variables used to classify the final states, and $\Delta\xi$ the size of the range of values around ξ that we decide to include in our counting. The *differential cross section* is defined as

$$\Delta\sigma_\alpha(\xi) = \frac{\Delta N_{\text{events}}(\alpha,\xi)}{\Delta t \Delta\xi N_t \Phi}\Delta\xi\,, \tag{2.4}$$

where now $\Delta N_{\text{events}}(\alpha,\xi)$ are the scattering events happening in the time interval t corresponding to the values α for the discrete variables and in an interval $\Delta\xi$ around ξ of the continuous variables. In the limit of infinitesimal time and ξ-intervals,[21]

$$\frac{d\sigma_\alpha(\xi)}{d\xi} = \frac{dN_{\text{events}}(\alpha,\xi)}{dt d\xi N_t \Phi}\,. \tag{2.5}$$

Collider experiments In collider experiments two beams of particles are directed at each other, usually circulating in opposite directions on the same circular trajectory. The beams are typically built out of several bunches of particles, each bunch having N_{b1} and N_{b2} particles in beam 1 and beam 2, respectively. In the period T that it takes to go around the whole circle (which we take to be the same for the two beams), two bunches will cross twice, so the frequency at which they collide is $2/T$. If the beams have the same cross-sectional area A_b and velocity v, over the time $T/2$ there will be $N_{b1}N_{b2}$ pairs of particles possibly interacting with each other

[21] A comment about the notation. The infinitesimal $d\xi$ in the denominator does not mean that one is taking derivatives. Rather, it indicates the variables according to which we are distributing our events. In other words, it indicates over what variables one should integrate to get integrated cross sections, e.g., corresponding to final states falling in *finite* intervals of the continuous variables. As an example, the differential cross section $\frac{d\sigma}{d^3p^{(1)}}(\vec{p}_1) = \frac{d\sigma}{dp_x^{(1)}dp_y^{(1)}dp_z^{(1)}}(\vec{p_1})$ for a two-body elastic scattering corresponds to the rate of events per unit flux per unit target with the momenta of particle 1 within an infinitesimal cube around $\vec{p_1}$. If one is to count events with $\vec{p_1}$ in a finite cube C around $\vec{p_1}$, then the relevant cross section is $\Delta\sigma = \int_C d^3p \frac{d\sigma}{d^3p^{(1)}}(\vec{p}_1)$.

when two bunches meet,[22] and so $\frac{\sigma}{A_b} N_{b1} N_{b2}$ scattering events. In each beam there is typically a large number of bunches, N_{B1} and N_{B2}, for each of which the same considerations apply. Following the same reasoning as above (and under the same conditions on the beam density), the number of events per unit time will then be

$$\frac{\Delta N_{\text{events}}}{\Delta t} = \frac{2}{T} \frac{N_{B1} N_{B2} N_{b1} N_{b2}}{A_b} \sigma = \mathcal{L} \sigma \,, \tag{2.6}$$

where $\mathcal{L}$ is the *luminosity* delivered by the collider. The analogue of Eq. (2.2) reads then

$$\sigma = \frac{1}{\mathcal{L}} \frac{\Delta N_{\text{events}}}{\Delta t} \,, \tag{2.7}$$

and similarly for the differential cross section.

Breit-Wigner formula The distinctive sign of an unstable particle being created as an intermediate state in a scattering process is the appearence of a peak in the cross section as a function of energy, which is called a *resonance*. The location and width of the peak are related to the mass and decay width of the unstable particle. A hand-waving way to see this is as follows: the state of an unstable system of approximate energy m created at t_0 and decaying exponentially with lifetime $1/\Gamma$ can be described by the wave function

$$\psi(t) = \psi(t_0) e^{-im(t-t_0)} e^{-\frac{\Gamma}{2}(t-t_0)} \theta(t-t_0) \,, \tag{2.8}$$

so that its magnitude decays as

$$|\psi(t)|^2 = |\psi(t_0)|^2 e^{-\Gamma(t-t_0)} \theta(t-t_0) \,. \tag{2.9}$$

Analysing the system in energy, which means taking the Fourier transform of Eq. (2.8), leads to

$$\tilde{\psi}(E) = \int dt\, e^{iEt} \psi(t) = \frac{i\psi(t_0)}{E - m + i\frac{\Gamma}{2}} e^{iEt_0} \,. \tag{2.10}$$

Assuming that such a state describes the intermediate stage of a scattering process, the total cross section $\sigma(E)$ will be proportional to the probability of observing this state at energy E, i.e.,

$$\sigma(E) \propto |\tilde{\psi}(E)|^2 \propto \frac{\left(\frac{\Gamma}{2}\right)^2}{(E-m)^2 + \left(\frac{\Gamma}{2}\right)^2} \,, \tag{2.11}$$

with the maximal value reached at $E = m$, and full width at half height equal to Γ. If a profile of this form is observed in experimental data, the position m of the peak and its width Γ are identified as the mass and decay width of an unstable particle (also referred to as a resonance). Equation (2.11) is known as the *Breit-Wigner formula*, and it provides a reasonable approximate description of many resonances.[23]

[22] We are neglecting here the loss of particles coming from the scattering events that actually take place: if the cross section of the process is small, so will be this loss, and it will be possible to neglect it for some time.

[23] The derivation given above is admittedly sketchy. The real reason why this functional form shows up is that in correspondence with a resonance one of the partial waves in the partial wave expansion of the cross section becomes maximal, which means that the related phase shifts passes through the value $\frac{\pi}{2}$ (see Ref. [3]).

Both scattering and decay processes are governed by the fundamental interactions, that tell us which processes can take place, and with what probability. The considerations made above about kinematical and dynamical constraints extend to the case of scattering processes: conservation laws put restrictions on the allowed processes, and such laws depend on the symmetries of the interactions. To understand how decays and scattering processes happen and which quantities they conserve, we have to look at how the various interactions work, but before we can do that we need to know how one relates theory with experiment.

2.2 Brief review of Quantum Mechanics

We begin with a brief recap of quantum mechanics. The quantum mechanical description of a physical system is in terms of vectors ψ (or rather rays $e^{i\omega}\psi$, with ω an arbitrary phase) of a Hilbert space $\mathcal{H}$, describing the state of the system by encoding the possible results and relative frequencies of experimental measurements. These are determined by the eigenvalues and eigenvectors of linear operators representing the physical observables.

Hilbert spaces A Hilbert space is a linear space, i.e., if $\psi_{1,2} \in \mathcal{H}$ then $\alpha_1\psi_1 + \alpha_2\psi_2 \in \mathcal{H}$ for any $\alpha_{1,2} \in \mathbb{C}$, endowed with a positive-definite scalar product (ϕ, ψ), obeying

$$(\phi, \alpha_1\psi_1 + \alpha_2\psi_2) = \alpha_1(\phi, \psi_1) + \alpha_2(\phi, \psi_2)\,, \qquad (\psi, \phi) = (\phi, \psi)^*\,, \qquad (\psi, \psi) \geq 0\,, \tag{2.12}$$

with the equal sign holding in the last inequality only if $\psi = 0$ is the zero vector. Associated with the scalar product is a norm $\|\psi\| \equiv (\psi, \psi)^{\frac{1}{2}}$. Any vector in $\mathcal{H}$ can be expressed as a linear combination of the elements of a complete orthonormal basis $\{e_n\}$, $n \in \mathbb{N}$, i.e., $\psi = \sum_n c_n e_n$, with $(e_n, e_m) = \delta_{nm}$,[24] and $\sum_n |c_n|^2 < \infty$.

Linear operators Linear operators A acting on $\mathcal{H}$ play a central role in Quantum Mechanics. A linear operator obeys

$$A(\alpha_1\psi_1 + \alpha_2\psi_2) = \alpha_1 A\psi_1 + \alpha_2 A\psi_2\,. \tag{2.13}$$

The adjoint $A^\dagger$ of an operator is defined by the relation

$$(\psi, A\phi) = (A^\dagger\psi, \phi)\,. \tag{2.14}$$

Eigenvectors $\psi_n^{(A)}$ of a linear operator are defined by the equation

$$A\psi_n^{(A)} = a_n\psi_n^{(A)}\,. \tag{2.15}$$

Unitary operators A *unitary operator* U is a linear operator defined by the properties (i) $\|U\psi\|^2 = (U\psi, U\psi) = (\psi, \psi) = \|\psi\|^2$, and (ii) the image of $\mathcal{H}$ under U is the whole Hilbert

[24] More precisely, a Hilbert space is a linear space with a scalar product which is furthermore complete with respect to the distance function $d(\psi, \phi) = \|\psi - \phi\|$. Completeness means that any Cauchy sequence of vectors ψ_n converges to an element $\psi \in \mathcal{H}$. A Cauchy sequence is one such that for every ϵ exists N such that $d(\psi_n, \psi_m) < \epsilon$ for all $n, m > N$. Hilbert spaces with a *countable* orthonormal basis are called *separable*.

space, $U\mathcal{H} = \mathcal{H}$. Property (i) is equivalent to $U^\dagger U = \mathbf{1}$,[25] while property (ii) is equivalent to $UU^\dagger = \mathbf{1}$,[26] which are not equivalent properties when the Hilbert space is infinite-dimensional.

In general, a unitary transformation corresponds to a change of basis in the Hilbert space, thus leading to an equivalent description of the physical system under consideration.

Observables Physical observables are represented in $\mathcal{H}$ by linear *Hermitean* operators $A = A^\dagger$. The eigenvectors of an operator of this type form a complete basis of $\mathcal{H}$, and the corresponding eigenvalues $a_n \in \mathbb{R}$ are real, corresponding to the possible experimental outcomes of a measurement. For a system in a state ψ, the probability of obtaining the result a_n for a measurement of A is given by

$$P_n^{(A)} = \frac{|(\psi_n^{(A)}, \psi)|^2}{(\psi_n^{(A)}, \psi_n^{(A)})(\psi, \psi)} \,. \tag{2.16}$$

The temporal evolution of a physical system is governed by its Hamiltonian $H = H^\dagger$, which is here assumed to be time-independent. The temporal evolution in the *Schrödinger picture* is expressed in terms of a time-dependent state vector $\psi(t) = e^{-iHt}\psi$, with $\psi = \psi(0)$ the state of the system at $t = 0$. The temporal evolution operator e^{-iHt} is a unitary operator. Linear operators corresponding to physical observables are time-independent in this picture, and so are their eigenvectors. If a measurement is made at time t, then in Eq. (2.16) the state vector $\psi(t)$ at time t must be used.

Dirac notation State vectors and their scalar products are often written using Dirac notation. State vectors ψ are represented by the *ket* $|\psi\rangle$, while the linear functional L_ψ, $L_\psi\phi \equiv (\psi, \phi)$, is represented by the *bra* $\langle\psi|$. Scalar products then read $(\phi, \psi) = \langle\phi|\psi\rangle$, and expectation values are written as $(\phi, A\psi) = \langle\phi|A|\psi\rangle$. While completely unambiguous for Hermitean operators, for a general operator it must be understood that $\langle\phi|A|\psi\rangle = \langle\phi|A\psi\rangle = \langle A^\dagger\phi|\psi\rangle$.

Schrödinger, Heisenberg, and Dirac (interaction) picture Quantum mechanics can be equivalently formulated in the Schrödinger picture described above, in which states evolve with time as determined by the Hamiltonian of the system, while observables are independent of time,

$$\text{Schrödinger picture:} \quad |\psi(t)\rangle_S = e^{-iHt}|\psi(0)\rangle_S \,, \qquad \mathcal{O}_S \,, \tag{2.17}$$

and in the *Heisenberg picture*, in which states are fixed at their $t = 0$ value while observables evolve with time as determined by the Hamiltonian of the system,

$$\text{Heisenberg picture:} \quad |\psi\rangle_H = |\psi(0)\rangle_S \,, \qquad \mathcal{O}_H(t) = e^{iHt}\mathcal{O}_S e^{-iHt} \,. \tag{2.18}$$

[25] This follows from norm preservation, $\|U\psi\| = \|\psi\|$, and the polarisation identity,

$$4(\psi, \phi) = \|\psi + \phi\| + \|\psi - \phi\| + i\|\psi - i\phi\| - i\|\psi + i\phi\| \,,$$

leading to

$$(U\psi, U\phi) = (\psi, \phi) = (\psi, U^\dagger U\phi) \,,$$

for arbitrary $\phi, \psi \in \mathcal{H}$.

[26] Since for every ψ there is a ϕ such that $\psi = U\phi$, then using associativity

$$\psi = U(U^\dagger U)\phi = (UU^\dagger)U\phi = (UU^\dagger)\psi \,.$$

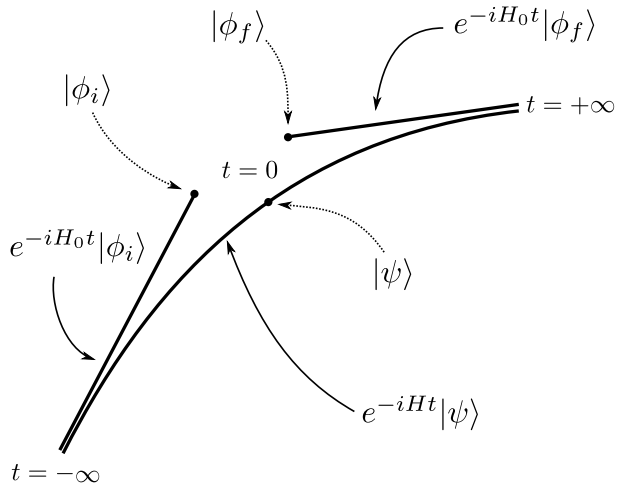

Figure 7: Schematic depiction of the relation between free and exact evolution of the system in a scattering process.

In this case the eigenvectors of $\mathcal{O}_H(t)$ are generally time-dependent, and if a measurement is made at time t, then in Eq. (2.16) the eigenvectors at time t must be used.

There is a third, intermediate picture, known as *interaction* or *Dirac picture*, where both the observables and the states evolve in time but in different ways. This picture is useful when the full Hamiltonian $H = H_0 + V$ can be split into a *free* Hamiltonian H_0, corresponding to a free system, and an interaction part V. In the interaction picture, observables obey the free temporal evolution determined by H_0, while the state vector evolves essentially with the interaction part only. More precisely,

$$\text{Dirac (interaction) picture:} \quad |\psi(t)\rangle_I = e^{iH_0t}e^{-iHt}|\psi(0)\rangle_S\,, \qquad \mathcal{O}_I(t) = e^{iH_0t}\mathcal{O}_S e^{-iH_0t}\,. \tag{2.19}$$

Notice that in general $e^{iH_0t}e^{-iHt} \neq e^{i(H_0-H)t} = e^{-iVt}$! Clearly, expectation values (and thus the physics) is independent of the picture one uses,

$$_S\langle\phi(t)|\mathcal{O}_S|\psi(t)\rangle_S = {}_H\langle\phi|\mathcal{O}_H(t)|\psi\rangle_H = {}_I\langle\phi(t)|\mathcal{O}_I(t)|\psi(t)\rangle_I\,, \tag{2.20}$$

as can be explicitly verified.

2.3 Formal theory of scattering

In and out states Let us discuss in more detail what it means that the initial and final states of the system in a scattering process look like freely-evolving particle states. In physical terms, this means that as $t \to -\infty$ or $t \to +\infty$ the exact state of the system, $e^{-iHt}|\psi\rangle$, which evolves with the full Hamiltonian, is practically indistinguishable from freely evolving states $e^{-iH_0t}|\phi_i\rangle$ and $e^{-iH_0t}|\phi_f\rangle$, respectively. Here H_0 denotes a free Hamiltonian, corresponding to a system of non-interacting particles. In mathematical terms, this translates into the fact that for states $|\psi\rangle$ describing a scattering system one has

$$\begin{aligned} \lim_{t\to-\infty} \|e^{-iHt}|\psi\rangle - e^{-iH_0t}|\phi_i\rangle\| = 0\,, \\ \lim_{t\to+\infty} \|e^{-iHt}|\psi\rangle - e^{-iH_0t}|\phi_f\rangle\| = 0\,, \end{aligned} \tag{2.21}$$

for certain $|\phi_{i,f}\rangle$. A pictorial representation of the situation is given in Fig. 7. Turning the argument around, this means that if we prepare our system in the distant past in the state $|\phi_i\rangle$, or more precisely if the evolution of the system that we prepare in the distant past looks for all practical purposes as $e^{-iH_0t}|\phi_i\rangle$, then the state vector that describes the exact temporal evolution of the system with the full Hamiltonian will be

$$|\psi_+\rangle = \lim_{t\to-\infty} e^{iHt}e^{-iH_0t}|\phi_i\rangle\,. \tag{2.22}$$

Similarly, if the state we observe in the distant future evolves for all practical purposes like $e^{-iH_0t}|\phi_f\rangle$, then its exact temporal evolution is described by the state vector[27]

$$|\psi_-\rangle = \lim_{t\to+\infty} e^{iHt}e^{-iH_0t}|\phi_f\rangle\,. \tag{2.23}$$

The states $|\psi_+\rangle$ and $|\psi_-\rangle$ are respectively the *in* and *out* states corresponding to $|\phi_i\rangle$ and $|\phi_f\rangle$, which we call instead the *asymptotic states* of the system.

Møller operators Equations (2.22) and (2.23) define two operators, the *scattering* or *Møller* operators,

$$\Omega_\pm = \lim_{t\to\mp\infty} e^{iHt}e^{-iH_0t}\,. \tag{2.24}$$

Being the limit of unitary operators, one can show that these operators conserve the norm, i.e., $\|\Omega_\pm|\phi\rangle\| = \||\phi\rangle\|$. Since the initial state can be prepared as we please, and anything that we want can be measured in the final state, $|\phi_i\rangle$ and $|\phi_f\rangle$ range over a complete set of states describing our system. From this and norm conservation one can conclude immediately that $\Omega_\pm^\dagger\Omega_\pm = \mathbf{1}$. In principle there might be states of the system that do not look like freely-evolving states as $t\to\mp\infty$, and which would therefore not be accessible in a scattering experiment: we will assume that this is not the case.[28] If so, then for all states $|\psi\rangle$ there are asymptotic initial and final states, i.e., the operators $\Omega_\pm$ map the whole Hilbert space into the whole Hilbert space, and are therefore unitary (see the discussion on page 26). This implies in particular that also $\Omega_\pm\Omega_\pm^\dagger = \mathbf{1}$ holds.

***S*-matrix** What we measure in experiments is not the exact temporal evolution of the system, which is inaccessible, but rather the transition probability for the initial state to be observed in some prescribed final state. If we have an initial state described by $|\psi_+(t)\rangle = e^{-iHt}|\psi_+\rangle \to e^{-iH_0t}|\phi_i\rangle$ as $t\to-\infty$, and at time T_f we project on the final state $|\psi_-(t)\rangle = e^{-iHt}|\psi_-\rangle$ which is such that $|\psi_-(t)\rangle \to e^{-iH_0t}|\phi_f\rangle$ as $t\to+\infty$, the relevant transition amplitude is given by

$$\langle\psi_-(T_f)|\psi(T_f)\rangle = \langle\psi_-|e^{iHT_f}e^{-iHT_f}|\psi_+\rangle = \langle\psi_-|\psi_+\rangle\,, \tag{2.25}$$

[27] An alternative viewpoint is that when me make a measurement on the system at time t we are projecting its state on some definite vector $e^{-iH_0t}|\phi_f\rangle$ corresponding to our experimental apparatus, which is then associated with a freely-evolving projector $e^{-iH_0t}|\phi_f\rangle\langle\phi_f|e^{+iH_0t}$. Its exactly-evolving counterpart is $e^{iHt}|\psi_+\rangle\langle\psi_+|e^{iHt}$, and the two projectors are the same in the limit $t\to+\infty$.

[28] In the non-relativistic case these would be the bound states of the system, like, e.g., a hydrogen atom state in *ep* scattering. In the relativistic setting the hydrogen atom counts as a particle, although not an elementary one, that can be used as an initial state or can be seen as a final state in a scattering process, so this does not constitute a problem.

which is time-independent. We can then compute it as follows:

$$\begin{aligned}\langle\psi_-|\psi_+\rangle &= \lim_{T_f\to+\infty}\langle\psi_-|e^{iHT_f}e^{-iHT_f}|\psi_+\rangle \\ &= \lim_{T_f\to+\infty,T_i\to-\infty}\langle\psi_-|e^{iHT_f}e^{-iHT_f}e^{iHT_i}e^{-iHT_i}|\psi_+\rangle \\ &= \lim_{T_f\to+\infty,T_i\to-\infty}\langle\phi_f|e^{iH_0T_f}e^{-iHT_f}e^{iHT_i}e^{-iH_0T_i}|\phi_i\rangle \\ &= \langle\phi_f|\Omega_-^\dagger\Omega_+|\phi_i\rangle \equiv \langle\phi_f|S|\phi_i\rangle = S_{fi}\,,\end{aligned} \tag{2.26}$$

where we have defined the *S-operator*, $S = \Omega_-^\dagger\Omega_+$, whose matrix elements S_{fi} constitute the *S-matrix*. More directly,

$$\langle\psi_-|\psi_+\rangle = \langle\phi_f|\Omega_-^\dagger\Omega_+|\phi_i\rangle = \langle\phi_f|S|\phi_i\rangle\,. \tag{2.27}$$

The S-matrix encodes all the relevant information about scattering processes: from the transition amplitudes S_{fi} one can get the transition *probabilities* $P_{fi} = |S_{fi}|^2$, which can be (indirectly) measured in experiments.

Properties of the S-matrix Let us discuss a few important properties of S. The operator S is *unitary*, being the product of unitary operators, so $S^\dagger S = SS^\dagger = \mathbf{1}$, respectively.[29] The physical meaning of unitarity is that probability is conserved: from an initial state we will certainly get to some final state, and a final state has certainly come from an initial state. These physically obvious statements correspond mathematically to $S^\dagger S = \mathbf{1}$ and $SS^\dagger = \mathbf{1}$. In fact, using the completeness relations $\sum_f |\phi_f\rangle\langle\phi_f| = \sum_i |\phi_i\rangle\langle\phi_i| = \mathbf{1}$, we find

$$\begin{aligned}\sum_f P_{fi} = \sum_f |S_{fi}|^2 = \sum_f \langle\phi_i|S^\dagger|\phi_f\rangle\langle\phi_f|S|\phi_i\rangle = \langle\phi_i|S^\dagger S|\phi_i\rangle\,, \\ \sum_i P_{fi} = \sum_i |S_{fi}|^2 = \sum_i \langle\phi_f|S|\phi_i\rangle\langle\phi_i|S^\dagger|\phi_f\rangle = \langle\phi_f|SS^\dagger|\phi_f\rangle\,.\end{aligned} \tag{2.28}$$

Next, notice that $\forall s$

$$\begin{aligned}e^{iHs}\Omega_\pm e^{-iH_0s} &= \lim_{t\to\mp\infty} e^{iHs}e^{iHt}e^{-iH_0t}e^{-iH_0s} = \lim_{t\to\mp\infty} e^{iH(t+s)}e^{-iH_0(t+s)} \\ &= \lim_{t\to\mp\infty} e^{iHt}e^{-iH_0t} = \Omega_\pm\,.\end{aligned} \tag{2.29}$$

Taking the derivative with respect to s of this relation and then setting s to zero we find the *intertwining relations*

$$H\Omega_\pm = \Omega_\pm H_0\,, \qquad H_0\Omega_\pm^\dagger = \Omega_\pm^\dagger H\,. \tag{2.30}$$

[29] Unitarity of S actually holds even if $\Omega_\pm$ are not unitary, but have as image only a subspace Λ of the whole Hilbert space, the same for both operators, corresponding to scattering states. In this case $\Omega_\pm\Omega_\pm^\dagger = \Pi_\Lambda$ equals the projector on this subspace. Nonetheless, since $\Pi_\Lambda\Omega_\pm = \Omega_\pm$ as Π_Λ acts like the identity on Λ, one has

$$\begin{aligned}S^\dagger S &= \Omega_+^\dagger\Omega_-\Omega_-^\dagger\Omega_+ = \Omega_+^\dagger\Pi_\Lambda\Omega_+ = \Omega_+^\dagger\Omega_+ = \mathbf{1}\,, \\ SS^\dagger &= \Omega_-^\dagger\Omega_+\Omega_+^\dagger\Omega_- = \Omega_-^\dagger\Pi_\Lambda\Omega_- = \Omega_-^\dagger\Omega_- = \mathbf{1}\,.\end{aligned}$$

It then follows that

$$H_0 S = H_0 \Omega_-^\dagger \Omega_+ = \Omega_-^\dagger H \Omega_+ = \Omega_-^\dagger \Omega_+ H_0 \Rightarrow [H_0, S] = 0\,. \tag{2.31}$$

If we now take the matrix element of this relation between initial and final states that are energy eigenstates we find

$$0 = \langle \phi_f | [H_0, S] | \phi_i \rangle = (E_f - E_i) \langle \phi_f | S | \phi_i \rangle = (E_f - E_i) S_{fi}\,, \tag{2.32}$$

which implies that S_{fi} can be nonzero only for initial and final states with the same energy. In other words, energy is conserved in a scattering process – as it should. Finally, if there is some symmetry generator G that commutes with both the free and the full Hamiltonians, $[G, H_0] = [G, H] = 0$, then it will commute with the scattering operators, $[G, \Omega_\pm] = 0$. In particular, interactions $V = H - H_0$ are usually translationally and rotationally invariant, so for the momentum $\vec{P}$ and the angular momentum $\vec{J}$ we have

$$[\vec{P}, \Omega_\pm] = [\vec{J}, \Omega_\pm] = 0\,. \tag{2.33}$$

From this it follows immediately that

$$[\vec{P}, S] = [\vec{J}, S] = 0\,, \tag{2.34}$$

i.e., momentum and angular momentum are conserved in scattering processes – again, as it should be. Lorentz invariance is somewhat trickier to achieve, but it is certainly desired. If we build V properly, then we will have $U_0(\Lambda)^\dagger S U_0(\Lambda) = S$, with $U_0(\Lambda)$ the unitary operator implementing the Lorentz transformation Λ on free particle states (see below Section 4).

Non-relativistic case For non-relativistic two-particle elastic scattering, factoring out the trivial behaviour of the centre-of-mass (CM) of the system, one is reduced to studying the relative motion in the CM system, which is equivalent to studying the scattering of one particle in a potential. Let the state of the system in the CM be described by the state vector $|\phi\rangle$. It is useful for certain applications to recast Eq. (2.24) as

$$\Omega_\pm = \mathbf{1} + \int_0^{\mp\infty} dt\, \frac{d}{dt} e^{iHt} e^{-iH_0 t} = \mathbf{1} + i \int_0^{\mp\infty} dt\, e^{iHt} V e^{-iH_0 t}\,, \tag{2.35}$$

where H_0, V, and H all involve only the relative coordinates and momenta of the system. Applied on $|\phi\rangle$, using the decomposition over momentum (improper) eigenstates $|\vec{p}\rangle$, one finds

$$\Omega_\pm |\phi\rangle = \int \frac{d^3p}{(2\pi)^3} \langle \vec{p} | \phi \rangle \left(|\vec{p}\rangle + i \int_0^{\mp\infty} dt\, e^{iHt} V e^{-iH_0 t} |\vec{p}\rangle \right)\,. \tag{2.36}$$

Here the normalisation is chosen to be $\langle \vec{p}\,' | \vec{p} \rangle = (2\pi)^3 \delta(\vec{p}\,' - \vec{p})$. Including a damping factor $e^{-\epsilon|t|}$ that will be eventually removed, and that should not change anything for well-defined states with a convergent integration over $\vec{p}$, one finds

$$\begin{aligned}
\Omega_\pm |\phi\rangle &= \lim_{\epsilon\to 0} \int \frac{d^3p}{(2\pi)^3} \langle \vec{p} | \phi \rangle \left(|\vec{p}\rangle + i \int_0^{\mp\infty} dt\, e^{i(H - E_{\vec{p}} \mp i\epsilon)t} V |\vec{p}\rangle \right) \\
&= \int \frac{d^3p}{(2\pi)^3} \langle \vec{p} | \phi \rangle \lim_{\epsilon\to 0} \left\{ \mathbf{1} + \frac{1}{E_{\vec{p}} - H \pm i\epsilon} V \right\} |\vec{p}\rangle \\
&= \int \frac{d^3p}{(2\pi)^3} \langle \vec{p} | \phi \rangle \lim_{\epsilon\to 0} \frac{\pm i\epsilon}{E_{\vec{p}} - H \pm i\epsilon} |\vec{p}\rangle \equiv \int \frac{d^3p}{(2\pi)^3} |\vec{p}\,\pm\rangle \langle \vec{p} | \phi \rangle\,.
\end{aligned} \tag{2.37}$$

The in and out state states $|\vec{p}\pm\rangle$ are eigenstates of momentum and of the (relative part of the) full Hamiltonian, $H|\vec{p}\pm\rangle = E_{\vec{p}}|\vec{p}\pm\rangle$ with the same momentum and the same energy $E_{\vec{p}}$ as $|\vec{p}\rangle$, although in that case this is the eigenvalue of the free Hamiltonian, $H_0|\vec{p}\rangle = E_{\vec{p}}|\vec{p}\rangle$. We write also

$$|\vec{p}\pm\rangle = \lim_{\epsilon\to 0} \frac{\pm i\epsilon}{E_{\vec{p}} - H \pm i\epsilon}|\vec{p}\rangle = \Omega_\pm(E_{\vec{p}})|\vec{p}\rangle\,, \tag{2.38}$$

which provides the definition of the energy-dependent scattering operators $\Omega_\pm(E_{\vec{p}})$. notice that

$$(E_{\vec{p}} - H)|\vec{p}\pm\rangle = \lim_{\epsilon\to 0} \frac{\pm i\epsilon(E_{\vec{p}} - H)}{E_{\vec{p}} - H \pm i\epsilon}|\vec{p}\rangle = 0\,, \tag{2.39}$$

i.e., $|\vec{p}\pm\rangle$ are eigenstates of H of energy $E_{\vec{p}} \geq 0$. Elastic scattering leads from a momentum eigenstate to another momentum eigenstate, so the relevant S-matrix elements read

$$\begin{aligned}
\langle\vec{p}\,' - |\vec{p}\,+\rangle &= \langle\vec{p}\,'|S|\vec{p}\rangle = \langle\vec{p}\,'|\Omega_-(E_{\vec{p}\,'})^\dagger\Omega_+(E_{\vec{p}})|\vec{p}\rangle \\
&= \langle\vec{p}\,'|\vec{p}\rangle + \langle\vec{p}\,'|\left[\Omega_-(E_{\vec{p}\,'}) - \Omega_+(E_{\vec{p}\,'})\right]^\dagger\Omega_+(E_{\vec{p}})|\vec{p}\rangle \\
&= \langle\vec{p}\,'|\vec{p}\rangle - \langle\vec{p}\,'|V\frac{2i\epsilon}{(H - E_{\vec{p}\,'})^2 + \epsilon^2}\Omega_+(E_{\vec{p}})|\vec{p}\rangle \\
&= \langle\vec{p}\,'|\vec{p}\rangle - \langle\vec{p}\,'|V\Omega_+(E_{\vec{p}})\frac{2i\epsilon}{(H_0 - E_{\vec{p}\,'})^2 + \epsilon^2}|\vec{p}\rangle \\
&= \langle\vec{p}\,'|\vec{p}\rangle - 2\pi i\delta(E_{\vec{p}} - E_{\vec{p}\,'})\langle\vec{p}\,'|V\Omega_+(E_{\vec{p}})|\vec{p}\rangle \\
&= \langle\vec{p}\,'|\vec{p}\rangle - 2\pi i\delta(E_{\vec{p}} - E_{\vec{p}\,'})\langle\vec{p}\,'|V|\vec{p}\,+\rangle \equiv \langle\vec{p}\,'|\vec{p}\rangle 2\pi i\delta(E_{\vec{p}} - E_{\vec{p}\,'})\frac{2\pi}{m}f(p,\theta,\varphi)\,.
\end{aligned} \tag{2.40}$$

The first term corresponds to no scattering, while the second one contains all the effects of the interaction.

One can connect the scattering amplitude f to the large-distance behaviour of scattering solutions of the Schrödinger equation. To see this, notice first the identity

$$\begin{aligned}
\frac{1}{E_{\vec{p}} - H_0 \pm i\epsilon}V\frac{1}{E_{\vec{p}} - H \pm i\epsilon} &= \frac{1}{E_{\vec{p}} - H_0 \pm i\epsilon}[(E_{\vec{p}} - H_0 \pm i\epsilon) - (E_{\vec{p}} - H \pm i\epsilon)]\frac{1}{E_{\vec{p}} - H \pm i\epsilon} \\
&= \frac{1}{E_{\vec{p}} - H \pm i\epsilon} - \frac{1}{E_{\vec{p}} - H_0 \pm i\epsilon}\,,
\end{aligned} \tag{2.41}$$

from which follows

$$\left(\mathbf{1} - \frac{1}{E_{\vec{p}} - H_0 \pm i\epsilon}V\right)|\vec{p}\pm\rangle = \left(\mathbf{1} - \frac{1}{E_{\vec{p}} - H_0 \pm i\epsilon}V\right)\left(\mathbf{1} + \frac{1}{E_{\vec{p}} - H \pm i\epsilon}V\right)|\vec{p}\rangle = |\vec{p}\rangle\,, \tag{2.42}$$

or equivalently

$$|\vec{p}\pm\rangle = |\vec{p}\rangle + \frac{1}{E_{\vec{p}} - H_0 \pm i\epsilon}V|\vec{p}\pm\rangle\,. \tag{2.43}$$

In the coordinate representation one finds

$$\psi_{p\pm}(\vec{x}) = \langle\vec{x}|\vec{p}\pm\rangle = e^{i\vec{p}\cdot\vec{x}} + \int d^3x'\,\langle\vec{x}|\frac{1}{E_{\vec{p}} - H_0 \pm i\epsilon}|\vec{x}\,'\rangle V(\vec{x}\,')\langle\vec{x}\,'|\vec{p}\pm\rangle\,. \tag{2.44}$$

Going over to polar coordinates and integrating over the angular variables we find

$$\begin{aligned}\langle\vec{x}\,|\frac{1}{E_p - H_0 \pm i\epsilon}|\vec{x}\,'\rangle &= \frac{2m}{(2\pi)^2 i}\frac{1}{|\vec{x}-\vec{x}\,'|}\int_0^\infty dp'\left(e^{ip'|\vec{x}-\vec{x}\,'|} - e^{-ip'|\vec{x}-\vec{x}\,'|}\right)\frac{p'}{p^2 - p'^2 \pm i\epsilon} \\ &= \frac{m}{2\pi^2 i}\frac{1}{|\vec{x}-\vec{x}\,'|}\int_{-\infty}^{+\infty} dp'\, e^{ip'|\vec{x}-\vec{x}\,'|}\frac{p'}{(p-p'\pm i\epsilon)(p+p'\pm i\epsilon)}\,.\end{aligned} \tag{2.45}$$

This integral can be computed using the residue theorem. Since $|\vec{x}-\vec{x}\,'|$ is positive, we can close the contour of integration in the upper half plane, thus picking the pole with positive imaginary part, i.e., $p' = \pm p + i\epsilon$. This finally yields

$$\langle\vec{x}\,|\frac{1}{E_p - H_0 \pm i\epsilon}|\vec{x}\,'\rangle = \frac{m}{2\pi^2 i}\frac{1}{|\vec{x}-\vec{x}\,'|}2\pi i e^{\pm ip|\vec{x}-\vec{x}\,'|}\frac{\mp p}{2p} = \mp\frac{m}{2\pi}\frac{1}{|\vec{x}-\vec{x}\,'|}e^{\pm ip|\vec{x}-\vec{x}\,'|}\,. \tag{2.46}$$

Plugging this into Eq. (2.44) we find

$$\psi_{p\pm}(\vec{x}) = e^{i\vec{p}\cdot\vec{x}} \mp \frac{m}{2\pi}\int d^3x'\,\frac{1}{|\vec{x}-\vec{x}\,'|}e^{\pm ip|\vec{x}-\vec{x}\,'|}V(\vec{x}\,')\psi_{p\pm}(\vec{x}\,')\,. \tag{2.47}$$

Let us focus now on $\psi_{p+}(\vec{x})$, and consider the limit of large $r \equiv |\vec{x}|$. For potentials with finite range (compact support, exponentials,...) this amounts to consider r much larger than such a range. Using

$$|\vec{x}-\vec{x}\,'| \simeq r - \hat{x}\cdot\vec{x}\,'\,, \tag{2.48}$$

where $\hat{x} = \vec{x}/r$, we obtain from Eq. (2.47)

$$\psi_{p+}(\vec{x}) = e^{i\vec{p}\cdot\vec{x}} - \frac{m}{2\pi}\frac{e^{ipr}}{r}\int d^3x'\,e^{-i\vec{p}\,'\hat{x}\cdot\vec{x}\,'}V(\vec{x}\,')\psi_{p+}(\vec{x}\,')\,, \tag{2.49}$$

where we have set $\vec{p}\,' = p\hat{x}$. But then

$$\begin{aligned}\frac{2\pi}{m}f(p,\theta,\varphi) &= -\langle\vec{p}\,'|V|\vec{p}+\rangle = -\int d^3x'\,\langle\vec{p}\,'|\vec{x}\,'\rangle V(\vec{x}\,')\langle\vec{x}\,'|\vec{p}+\rangle \\ &= -\int d^3x'\,e^{-i\vec{p}\,'\cdot\vec{x}\,'}V(\vec{x}\,')\psi_{p+}(\vec{x}\,')\,,\end{aligned} \tag{2.50}$$

and so comparing the two equations we see that

$$\psi_{p+}(\vec{x}) = e^{i\vec{p}\cdot\vec{x}} + \frac{e^{ipr}}{r}f(p,\theta,\varphi)\,. \tag{2.51}$$

The main conclusion is that the scattering amplitude (and therefore, as we will see, the differential cross section measured in experiments) is determined by the asymptotic, large-distance behaviour of the positive-energy solutions of the Schrödinger equation satisfying the following boundary condition: at large distances it looks like the sum of a plane wave, corresponding to the incoming particle, and an outgoing spherical wave. The relative weight of the two terms determines the probability of detecting the particle scattered in a given direction.

Dyson's formula The calculation above applies to the NR case, and can be used directly in a limited amount of cases where the exact solution of the time-independent Schrödinger equation can be obtained analytically. For more general cases one needs a way to obtain approximate results for the S-matrix elements that do not require the exact solution of the eigenvalue problem. To this end, we now work out a useful formula for the S operator.

Starting from its definition, we can write

$$S = \Omega_-^\dagger \Omega_+ = \lim_{t_2\to+\infty} \lim_{t_1\to-\infty} e^{iH_0 t_2} e^{-iH t_2} e^{iH t_1} e^{-iH_0 t_1} = \lim_{t_2\to+\infty} \lim_{t_1\to-\infty} \mathcal{U}(t_2, t_1)\,, \tag{2.52}$$

where $\mathcal{U}(t_2, t_1)$ is a unitary operator. To obtain an explicit expression for it, we will write down the differential equation that it obeys, and solve it subject to the initial condition $\mathcal{U}(t, t) = \mathbf{1}$. Taking the derivative with respect to t_2

$$\begin{aligned} \frac{\partial}{\partial t_2}\mathcal{U}(t_2, t_1) &= e^{iH_0 t_2} i(H_0 - H) e^{-iH_0 t_2} \mathcal{U}(t_2, t_1) = -i e^{iH_0 t_2} V e^{-iH_0 t_2} \mathcal{U}(t_2, t_1) \\ &= -i V_I(t_2) \mathcal{U}(t_2, t_1)\,, \end{aligned} \tag{2.53}$$

where

$$V_I(t) \equiv e^{iH_0 t} V e^{-iH_0 t}\,. \tag{2.54}$$

Notice that the temporal evolution is governed by the free Hamiltonian rather than the full one: this means that $V_I(t)$ is the interaction part of the Hamiltonian in the interaction picture (see discussion on page 28), which differs from the interaction Hamiltonian in the Heisenberg picture, $V(t) = e^{iHt} V e^{-iHt}$. Since $\mathcal{U}(t_2, t_1)^\dagger = \mathcal{U}(t_1, t_2)$, taking the derivative with respect to t_1 will not teach us anything new. For completeness, we report the result:

$$\frac{\partial}{\partial t_1}\mathcal{U}(t_2, t_1) = \mathcal{U}(t_2, t_1) i V_I(t_1)\,. \tag{2.55}$$

The solution of Eqs. (2.53) and (2.55) with the prescribed initial condition is

$$\begin{aligned} \mathcal{U}(t_2, t_1) &= \mathrm{Texp}\left\{-i\int_{t_1}^{t_2} dt\, V_I(t)\right\} \\ &= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int_{t_1}^{t_2} d\tau_1 \ldots \int_{t_1}^{t_2} d\tau_n\, T\left\{V_I(\tau_1)\ldots V_I(\tau_n)\right\}\,, \end{aligned} \tag{2.56}$$

where the *time-ordering symbol* T places the operators in descending order with respect to time starting from the left:[30]

$$\begin{aligned} T\{A_1(t_1)A_2(t_2)\} &= \theta(t_1 - t_2) A_1(t_1) A_2(t_2) + \theta(t_2 - t_1) A_2(t_2) A_1(t_1)\,, \\ T\{A_1(t_1)\ldots A_n(t_n)\} &= \sum_P \theta(t_{P(1)} - t_{P(2)}) \ldots \theta(t_{P(n-1)} - t_{P(n)}) A_{P(1)}(t_{P(1)}) \ldots A_{P(n)}(t_{P(n)})\,, \end{aligned} \tag{2.57}$$

where the sum is over all the distinct permutations P of $\{1, \ldots, n\}$. To see that the expression in Eq. (2.56) is indeed the solution of our equation, we rewrite it identically as follows,

$$\mathcal{U}(t_2, t_1) = \sum_{n=0}^{\infty} (-i)^n \int_{t_1}^{t_2} d\tau_1 \int_{t_1}^{\tau_1} d\tau_2 \ldots \int_{t_1}^{\tau_{n-1}} d\tau_n\, V_I(\tau_1)\ldots V_I(\tau_n)\,, \tag{2.58}$$

[30] The term with $n = 0$ is simply the identity operator.

where we have used the fact that for a given n all the $n!$ permutations of the times τ_j give the same contribution. We then see that t_2 appears only as the integration limit of the leftmost integral, and we then find straightforwardly that

$$\begin{aligned}\frac{\partial}{\partial t_2}\mathcal{U}(t_2,t_1) &= -iV_I(t_2)\sum_{n=1}^{\infty}(-i)^{n-1}\int_{t_1}^{t_2}d\tau_2\ldots\int_{t_1}^{\tau_{n-1}}d\tau_n\,V_I(\tau_2)\ldots V_I(\tau_n)\\ &= -iV_I(t_2)\sum_{n=0}^{\infty}(-i)^{n}\int_{t_1}^{t_2}d\tau_1\ldots\int_{t_1}^{\tau_{n-1}}d\tau_n\,V_I(\tau_1)\ldots V_I(\tau_n)\\ &= -iV_I(t_2)\mathcal{U}(t_2,t_1)\,.\end{aligned} \tag{2.59}$$

Since it is obvious that $\mathcal{U}$ in Eq. (2.58) satisfies the initial condition, it is the unique solution of our problem (as per the uniqueness of the solution to a Cauchy problem). In terms of $\mathcal{U}$ we can then write *Dyson's formula* for the S-operator:

$$S=\mathcal{U}(+\infty,-\infty)=\mathrm{Texp}\left\{-i\int_{-\infty}^{+\infty}dt\,V_I(t)\right\}\,. \tag{2.60}$$

This expression does not lead us any closer to solving exactly the scattering problem, but allows for a straightforward approximation scheme. If V is a small perturbation, then it makes sense to expand the time-ordered exponential in powers of the interaction, and compute the contribution of the first few terms: this should provide a reasonable approximation of the S-matrix elements. One then only needs to supply a suitable interaction term V in order to obtain physical predictions. In order to connect these to experiments, though, we still need to connect the S-matrix with quantities that can actually be measured, like the cross section of a process.

2.4 Cross sections from the S-matrix

The cross section of a process is directly related to the transition probability P_{fi} from an initial state $|\phi_i\rangle$ to a final state $|\phi_f\rangle$, and in fact it is the closest thing to P_{fi} that we can measure. There are in fact practical limitations to measure P_{fi} directly, most prominently the fact that the initial state in one single given scattering process is not known with arbitrary accuracy: the particle states used in scattering experiments are obtained through practical processes (e.g., acceleration of particles, preparation of beams) that are affected by inherent uncertainties, which do not allow for an exceedingly accurate determination of the actual wave function of the state. From a practical point of view, what can be measured is the transition probability averaged over many experiments, corresponding to many slightly different initial states. Luckily enough, if the experiment is designed with sufficient care and the initial states are sufficiently peaked around definite momenta of the particles, then the variation over the initial state turns out to have no effect, and what gets actually measured is the transition probability between idealised initial and final momentum eigenstates. We will not discuss the details here, which requires a detailed calculation using a wave-packet description of the initial and final states and a careful consideration of how scattering experiments are carried out. We simply reassure the reader that this can be done, reaching the same conclusions that will be obtained below with a simpler method. The upshot is that if we have a theory from which we can compute S-matrix elements, then these can be used to predict the outcome of scattering experiments, so allowing one to test the theory.

Recall that differential cross sections are operatively defined through the formula

$$\Delta\sigma_\alpha(\xi) = \frac{1}{N_t\Phi}\frac{\Delta N_{\text{events}}(\alpha,\xi)}{\Delta t\Delta\xi}\Delta\xi\,, \tag{2.61}$$

which directly applies to fixed-target experiments. Here $\Delta N_{\text{events}}(\xi)$ is the number of scattering events taking place in the time interval Δt and characterised by specific values of certain discrete quantum numbers, collectively denoted by α (particle type, charge, spin component), and by continuous quantum numbers ξ_j, collectively denoted by ξ, lying within the intervals $[\xi_j, \xi_j+\Delta\xi_j]$, with $\Delta\xi = \prod_j \Delta\xi_j$. Together, α and ξ characterise the final state of the system. Moreover, N_t is the number of scatterers in that part of the target where the beam impinges, and Φ is the beam flux. Together with the type of particles involved, these characterise the initial state of the system. In the following we will simplyfy the notation and indicate the specific combination of initial and final state by the subscript *fi*.

Consider now an idealised process in which a single particle from the beam (particle 1) possibly interacts with a single particle in the target (particle 2), corresponding to $N_t = N_b = 1$. For sufficiently diluted beams and targets, this is an accurate description of the experimental situation. Then, over the time Δt that it takes for the collision to complete, in a volume ΔV_b around the beam particle one will find only particles 1 and 2. This volume is just the inverse of the beam spatial density $\rho_b = N_b/V_b$ (where V_b is the total volume occupied by the beam), since $\rho_b\Delta V_b = 1$ means that a single particle is typically found in ΔV_b. Therefore, if particle 1 has velocity v then the beam flux equals $\Phi = \rho_b v = (1/\Delta V_b)v$. The average number of scattering events ΔN_{events} in this case is the probability that an interaction actually takes place leading form the initial state i to the specified final state f, i.e., the transition probability P_{fi},

$$P_{fi} = \frac{S_{fi}}{\langle\phi_f|\phi_f\rangle\langle\phi_i|\phi_i\rangle} = \frac{|\langle\phi_f|S|\phi_i\rangle|^2}{\langle\phi_f|\phi_f\rangle\langle\phi_i|\phi_i\rangle}\,. \tag{2.62}$$

While very small on the macroscopic scales experienced by the experimenters, the time Δt and the volume ΔV_b are very large on the microscopic scale experienced by the particles, determined by the typical length and time scale of the collision process. One can then eventually take the limits $\Delta t, \Delta V_b \to \infty$ without any appreciable loss of precision.

The next step in our idealisation is to consider particles with sharply defined momenta in the initial state, and to measure with infinite precision the momenta of the final state, so studying the transition from a momentum eigenstate to another momentum eigenstate. However, here we have to pause and consider carefully what we are doing. If we take the infinite-time and infinite-volume limit too soon, in fact, we run into problems. In an infinite spatial volume the momentum eigenstates are improper, non-normalisable eigenstates, for which Eq. (2.62) makes no sense. Furthermore, since S commutes with energy and momentum, its matrix elements between momentum eigenstates can certainly be written as

$$S_{fi} = \delta_{fi} + i(2\pi)^4\delta^{(4)}(P_f - P_i)\mathcal{M}_{fi}\,, \tag{2.63}$$

where the term $\delta_{fi} = \langle\phi_f|\phi_i\rangle$, corresponding to no scattering actually taking place, has been singled out, and $P_{i,f}$ are the total initial and final momentum. In the case of different initial and final states ($f \neq i$) that we will be considering here, only the second term matters. One sees immediately that the numerator in Eq. (2.62) contains the square of a Dirac delta, which again makes no sense. On the other hand, there was no reason to take the infinite-volume limit

immediately: that is certainly a sensible thing to do, as explained above, but one should do it with care.

Let us take advantage of the existence of a finite spacetime box in which the scattering process is essentially taking place. Consider a $T \times V = T \times L^3$ box of temporal extension T, which will be identified with Δt, and of spatial extension L and volume $V = L^3$, which will be identified with ΔV_b. The use of capital letters is to remind ourselves that these microscopic scales are actually huge scale for the particles, and will eventually be sent to infinity. For this reason, it should make no difference if we artificially impose periodic boundary conditions on our box, which will play no role in the infinite-box limit, but guarantee us translational invariance. In this setup, momentum eigenvalues are quantised, $\vec{p} = \frac{2\pi}{L}\vec{n}$ with $n_j \in \mathbb{Z}$, and one finds a single mode in a small momentum-space box of volume $\Delta^3 p = \frac{(2\pi)^3}{V}$. In a finite volume momentum eigenmodes are normalisble, and we choose the normalisation ${}_V\langle\vec{p}'|\vec{p}\rangle_V = 2p^0 V \delta^{(3)}_{\vec{p}',\vec{p}}$, where the delta is here a Kronecker delta. The reason for doing this is that in the infinite-volume limit one finds

$$1 = \sum_{\vec{p}} \frac{1}{2p^0 V} {}_V\langle\vec{p}'|\vec{p}\rangle_V \underset{V\to\infty}{\to} \int \frac{d^3p}{(2\pi)^3 2p^0} \left(\lim_{V\to\infty} {}_V\langle\vec{p}'|\vec{p}\rangle_V \right) = \int \frac{d^3p}{(2\pi)^3 2p^0} \langle\vec{p}'|\vec{p}\rangle \,, \tag{2.64}$$

meaning that

$$\lim_{L\to\infty} L\delta_{p'_j,p_j} = 2\pi\delta(p'_j - p_j) \,, \tag{2.65}$$

and that

$$\langle\vec{p}'|\vec{p}\rangle = (2\pi)^3 2p^0 \delta^{(3)}(\vec{p}' - \vec{p}) \,. \tag{2.66}$$

One then finds an acceptable normalisation for the infinite-volume momentum eigenstates, which is furthermore Lorentz-invariant (this can be seen directly, or simply noticing that since the integration measure is invariant, so is the normalisation – see Section A.5). Equation (2.66) is known as the *relativistic invariant normalisation* of momentum eigenstates. Translation invariance imposes now that the finite-box version of the S-operator reads

$$S^{(T,V)}_{fi} = \delta^{(T,V)}_{fi} + iTV\delta^{(4)}_{P_f,P_i}\mathcal{M}^{(T,V)}_{fi} \,, \tag{2.67}$$

having factored out TV so that $TV\delta^{(4)}_{P_f,P_i} \to (2\pi)^4\delta^{(4)}(P_f - P_i)$, and so $\mathcal{M}^{(T,V)}_{fi} \to \mathcal{M}_{fi}$ becomes its infinite-volume counterpart appearing in Eq. (2.63) in the infinite-box limit.

We are now ready to derive the relation between the S-matrix and the cross section. We consider a generic final state (as long as $f \neq i$), with n particles of momenta $\vec{p}'_j$. As explained above, in the setting considered here Eq. (2.61) becomes

$$\Delta\sigma = \frac{P_{fi}}{\frac{T}{V}v} \,, \tag{2.68}$$

having set $\Delta N_{\text{events}} = P_{fi}$, $N_t = N_b = 1$, $\Delta t = T$, and $\Delta V_b = V$. Here $\Delta\sigma$ indicates that we are computing a (discretised) differential cross section. Using Eq. (2.67), we find

$$\Delta\sigma = \frac{V}{Tv} \frac{(TV)^2\delta^{(4)}_{P_f,P_i}|\mathcal{M}^{(T,V)}_{fi}|^2}{4p^0_1 V p^0_2 V} \prod_{j=1}^{n} \frac{1}{2p'^0_j V} = \frac{TV\delta^{(4)}_{P_f,P_i}|\mathcal{M}^{(T,V)}_{fi}|^2}{4p^0_1 p^0_2 v} \prod_{j=1}^{n} \frac{\Delta^3 p'_j}{(2\pi)^3 2p'^0_j} \,, \tag{2.69}$$

where again $\Delta^3 p'_j = (2\pi)^3/V$ is the size of the bin in which the momentum of final particle j falls, that will eventually become the infinitesimal momentum-space element $d^3p'_j$ in the infinite-volume limit. Squaring the Kronecker delta causes no problems, and the factors of T and V cancel out to leave just what is required for a single regularised Dirac delta. We can now take $T, V \to \infty$ and find $\Delta\sigma \to d\sigma$ with

$$d\sigma = \frac{|\mathcal{M}_{fi}|^2}{4p_1^0 p_2^0 v}(2\pi)^4\delta^{(4)}(P_f - P_i)\prod_{j=1}^{n}\frac{d^3p'_j}{(2\pi)^3 2p'^0_j} = \frac{|\mathcal{M}_{fi}|^2}{4p_1^0 p_2^0 v}d\Phi^{(n)}\,, \tag{2.70}$$

where $d\Phi^{(n)}$ is the infinitesimal invariant element of the n-particle phase space $d\Phi^{(n)}$ (see Section A.5). This formula relates theory and experiment in the case of fixed-target experiments, with all quantities measured in the laboratory frame. On the other hand, $d\sigma$ is a Lorentz-invariant concept, expressing the "effective area" surrounding a target particle, as seen by a beam particle that impinges on it perpendicularly: being transverse to the motion of the beam particle, it is left invariant by a boost in the longitudinal direction of the beam.[31] Moreover, $d\Phi^{(n)}$ is Lorentz-invariant, and so is $\mathcal{M}_{fi}$ if we obtain it from a Lorentz-invariant theory and use a relativistic-invariant normalisation for the momentum eigenstates, as we are doing here. The only non-invariant piece in Eq. (2.70) is the denominator, which is specific to the fixed-target setup, and reads more explicitly $p_1^0 p_2^0 v = E_1 m_2 v = |\vec{p}_1| m_2$. This equals a suitable Lorentz-invariant quantity evaluated in the lab frame, namely

$$|\vec{p}_1| m_2 = \sqrt{\vec{p}_1^{\,2} m_2^2} = \sqrt{(E_1^2 - m_1^2)m_2^2} = \sqrt{(p_1\cdot p_2)^2 - p_1^2 p_2^2} \equiv I\,. \tag{2.71}$$

We can then write the formula for the differential cross section in a manifestly Lorentz-invariant way as follows,

$$d\sigma = \frac{|\mathcal{M}_{fi}|^2}{4I}d\Phi^{(n)}\,. \tag{2.72}$$

This is the formula we were after.

To make further progress, a detailed theory to compute $\mathcal{M}_{fi}$ is needed. This requires constructing an interaction Hamiltonian compatible with Lorentz invariance and the other symmetries of nature. In the next Section we will briefly describe how this can be done.

3 From Relativistic Quantum Mechanics to Quantum Field Theory

3.1 Relativistic quantum mechanics

3.1.1 Klein-Gordon-Schrödinger equation

In the nonrelativistic, free-particle case the fundamental dynamical equation, i.e., the Schrödinger equation,

$$i\partial_t\psi(t,\vec{x}) = -\frac{\vec{\nabla}^{\,2}}{2m}\psi(t,\vec{x})\,, \tag{3.1}$$

[31] Defining the flux using the projection of the beam velocity on the perpendicular of the target surface, $d\sigma$ remains invariant also under general Lorentz transformations.

obeyed by the particle wave function $\psi(t,\vec{x})$, can be derived from the relation between energy and momentum by upgrading the classical variables E and $\vec{p}$ to differential operators acting on the particle's wave function. In fact, starting from the non-relativistic energy-momentum relation

$$E = \frac{\vec{p}^{\,2}}{2m}\,, \tag{3.2}$$

if one replaces $E \to i\partial_t$ and $\vec{p} \to -i\vec{\nabla}$, and requires that Eq. (3.2) holds as a relation between differential operators acting on ψ, one recovers Eq. (3.1). It is interesting to notice that the differential operators associated with energy and momentum generate translations in time and space, respectively.[32]

The next thing to do to is to make the dynamics compatible with special relativity. In the relativistic case the energy-momentum relation is $E^2 - \vec{p}^{\,2} = m^2$ which after the same substitutions for E and $\vec{p}$ as above leads to the *Klein-Gordon equation*,

$$(\Box + m^2)\psi(t,\vec{x}) = (\partial_t^2 - \vec{\nabla}^2 + m^2)\psi(t,\vec{x}) = 0\,. \tag{3.3}$$

Schrödinger did not live under a stone and was perfectly aware that physics had to comply with special relativity, and this was in fact the first equation he wrote, before Klein and Gordon. However, he was forced to content himself with its non-relativistic limit, Eq. (3.1), which provides a consistent description of a particle in terms of a wave function ψ encoding the probability to find it somewhere in space at a given time, while one faces a number of technical difficulties in attempting a similar interpretation for the solutions of the fully relativistic equation, Eq. (3.3).

Negative probabilities and negative energy states The first difficulty is the lack of a covariant probability current giving a positive-definite probability density. From the Klein-Gordon equation and its complex conjugate one finds

$$0 = \phi^*(\Box + m^2)\phi - \phi(\Box + m^2)\phi^* = \phi^*\partial_\mu\partial^\mu\phi - \phi\partial_\mu\partial^\mu\phi^* = \partial_\mu[\phi^*\partial^\mu\phi - (\partial^\mu\phi^*)\phi]\,, \tag{3.4}$$

and so the current

$$J^\mu = i\phi^*\overset{\leftrightarrow}{\partial^\mu}\phi\,, \qquad f\overset{\leftrightarrow}{\partial^\mu}g \equiv f\partial^\mu g - (\partial^\mu f)g\,, \tag{3.5}$$

is conserved, i.e., $\partial_\mu J^\mu = 0$. This is therefore a good candidate for a probability current, since it is a four-vector (so correct transformation properties under Lorentz transformations) and it is conserved (so obeying a continuity equation, i.e., "probability is conserved"), but unfortunately its $\mu = 0$ component, i.e., the probability density, is not positive-definite.

The second difficulty is the existence of negative-energy solutions. It is not hard to see that the general solution of Eq. (3.3) is a superposition of plane waves, $\phi_p(x) = e^{-ip\cdot x}$, with p^μ an arbitrary four-vector satisfying the mass-shell condition $(p^0)^2 - \vec{p}^{\,2} = m^2$, corresponding to the four-momentum of the solution. In particular, the energy of the solution is read off from

[32] Derivatives "generating translations" can be understood as follows. The translated function $f_a(s) \equiv f(s+a)$ is obtained from the function $f(s)$ by adding up (infinitely many) translations by an infinitesimal amount ϵ, until the desired finite amount a is reached. For infinitesimal ϵ one has $f_\epsilon(s) = f(s) + \epsilon\partial_s f(s) + \mathcal{O}(\epsilon^2)$, and it is clear that the effect of such a translation is entirely encoded in $\partial_s f(s)$, i.e., in the effect of the derivative operator. To all orders,

$$f(s+a) = \sum_{n=0}^{\infty}\frac{a^n}{n!}\partial_s^n f(0) = e^{a\partial_s}f(s) = e^{ia(-i\partial_s)}f(s)\,.$$

$E\phi_p(x) = i\partial_0\phi_p(x) = p^0\phi_p(x)$, and since only $|p^0|$ is determined by the mass-shell condition, states with unbounded negative energy are present. This is in contradiction with the observed stability of matter: if such negative-energy states existed in nature, then nothing would prevent matter from decaying into states of more and more negative energy.

General solution of the Klein-Gordon equation Despite these difficulties, let us write the most general solution of the Klein-Gordon equation. The simplest way to do so is by going over to momentum space by a Fourier transform,

$$\phi(x) = \int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x}\tilde{\phi}(p)\,, \tag{3.6}$$

and observe that $\tilde{\phi}(p)$ obeys an algebraic equation,

$$\begin{aligned} 0 = (\Box + m^2)\phi(x) &= \int \frac{d^4p}{(2\pi)^4}\,[(\Box + m^2)e^{-ip\cdot x}]\tilde{\phi}(p) = \int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x}(-p^2 + m^2)\tilde{\phi}(p) \\ &\Rightarrow (p^2 - m^2)\tilde{\phi}(p) = 0\,. \end{aligned} \tag{3.7}$$

This restricts $\tilde{\phi}(p)$ to be non-vanishing only on the mass shell $p^2 = (p^0)^2 - \vec{p}^{\,2} = m^2$, which can be implemented by setting $\tilde{\phi}(p) = 2\pi\delta(p^2 - m^2)\tilde{\varphi}(p^0, \vec{p})$. Using the properties of the Dirac delta this implies [see Eq. (A.79)]

$$\begin{aligned} \phi(x) &= \int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x}\frac{2\pi}{2\sqrt{\vec{p}^{\,2} + m^2}}\left(\delta(p^0 - \sqrt{\vec{p}^{\,2} + m^2}) + \delta(p^0 + \sqrt{\vec{p}^{\,2} + m^2})\right)\tilde{\varphi}(p^0, \vec{p}) \\ &= \int \frac{d^3p}{(2\pi)^3 2p^0}\left(\tilde{\varphi}(p^0, \vec{p})e^{-i(p^0x^0 - \vec{p}\cdot\vec{x})} + \tilde{\varphi}(-p^0, \vec{p})e^{-i(-p^0x^0 - \vec{p}\cdot\vec{x})}\right)\Big|_{p^0 = \sqrt{\vec{p}^{\,2} + m^2}} \\ &= \int \frac{d^3p}{(2\pi)^3 2p^0}\left\{a(\vec{p})e^{-ip\cdot x} + b(\vec{p})^* e^{ip\cdot x}\right\} = \int d\Omega_p \left\{a(\vec{p})e^{-ip\cdot x} + b(\vec{p})^* e^{ip\cdot x}\right\}\,, \end{aligned} \tag{3.8}$$

where p^0 is now fixed by the mass-shell condition and restricted for convenience to take positive values only, $p^0 = \sqrt{\vec{p}^{\,2} + m^2}$, we set $a(\vec{p}) = \tilde{\varphi}(p^0, \vec{p})$ and $b(\vec{p})^* = \tilde{\varphi}(-p^0, -\vec{p})$, and $d\Omega_p$ denotes the invariant one-particle phase-space element [see Eq. (A.82)]. The first and second term in braces correspond respectively to positive- and negative-energy solutions of spatial momentum $\vec{p}$ and $-\vec{p}$, multiplied by (generally complex) amplitudes $a(\vec{p})$ and $b(\vec{p})^*$. With the chosen normalisation of the coefficients, the spacetime integral of J^0 reads

$$Q = \int d^3x\, J^0(x) = \int d\Omega_p\,[a(\vec{p})^* a(\vec{p}) - b(\vec{p})b(\vec{p})^*]\,. \tag{3.9}$$

It is clear that Q can take negative values, but it is interesting to notice that negative contributions come only from the negative-energy normal modes.

3.1.2 Dirac equation

Being the transposition into quantum-mechanical language of the relativistic energy-momentum relation $E^2 = \vec{p}^{\,2} + m^2$, the Klein-Gordon equation should be obeyed independently of the type of particle, in particular of its spin. On the other hand, there is nothing in it that specifies the

number of possible spin states, and so one should look for more restrictive equations if one wants to include spin in the description.

The appropriate equation for spin-$\frac{1}{2}$ fermions was found by Dirac, in an attempt to solve the problems of the Klein-Gordon equation mentioned above, i.e., negative energy states, and lack of a covariant probability current giving a positive-definite probability density. Negative-energy states appear because the equation is second-order in time and so, Dirac reasoned, a way to get rid of them would be to find a relativistic equation for the fermions that is first order in the time derivative. Lorentz invariance then requires that it be first order also in the spatial derivatives. Moreover, this equation should imply the Klein-Gordon equation. Dirac's idea was then to look for the "square root" of the Klein-Gordon equation, i.e., for an equation of the form

$$ib^\mu \partial_\mu \psi = a\psi \,, \tag{3.10}$$

satisfied by the wave function ψ, which implies automatically

$$-\, b^\mu \partial_\mu b^\nu \partial_\nu \psi = a^2 \psi \Longrightarrow (b^\mu \partial_\mu b^\nu \partial_\nu + a^2)\psi = 0 \,, \tag{3.11}$$

and to impose that this reproduces the Klein-Gordon equation. This requirement is satisfied if

$$\begin{aligned} \Box = b^\mu b^\nu \partial_\mu \partial_\nu = \tfrac{1}{2}\{b^\mu, b^\nu\}\partial_\mu \partial_\nu \,, \\ m^2 = a^2 \,, \end{aligned} \tag{3.12}$$

which in turn will be satisfied if

$$\{b^\mu, b^\nu\} = 2\eta^{\mu\nu} \,, \qquad a^2 = m^2 \,. \tag{3.13}$$

These equations cannot be solved by means of complex constants: while the second equation is easily solved by setting $a = \pm m$, the first one in the case $\mu = \nu$ would imply that $(b^\mu)^2 = \pm 1$, which is incompatible with $b^\mu b^\nu = 0$ for $\mu \neq \nu$. The simplest way to solve Eq. (3.13) is by means of 4×4 matrices, $b^\mu = \gamma^\mu$, $a = m\mathbf{1}_4$, with

$$\gamma^0 = \begin{pmatrix} \mathbf{1}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & -\mathbf{1}_2 \end{pmatrix} \,, \qquad \gamma^j = \begin{pmatrix} \mathbf{0}_2 & \sigma_j \\ -\sigma_j & \mathbf{0}_2 \end{pmatrix} \,. \tag{3.14}$$

One can verify explicitly that the γ^μ satisfy the relation[33]

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \,. \tag{3.15}$$

This means that ψ is upgraded to a four-component object, called *bispinor*,

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \,, \tag{3.16}$$

where $\psi_{1,2}$ are two-component objects called *spinors*. The resulting equation,

$$(i\partial\!\!\!/ - m)\psi = 0 \,, \tag{3.17}$$

[33]The matrices in Eq. (3.14) are not the only ones satisfying this relation, e.g., any set $\gamma^\mu_U = U^\dagger \gamma^\mu U$ obtained by a unitary transformation still does. A set of matrices satisfying Eq. (3.15) is said to generate a representation of the *Clifford algebra* associated with the metric defined by the Minkowski tensor $\eta^{\mu\nu}$.

where $\not{\partial} \equiv \gamma^\mu \partial_\mu$, is the *Dirac equation*. This equation solved the problem of the non-positive probability density, but as we will see below it did not solve the problem of negative energies. In fact, taking the Hermitean conjugate of Eq. (3.17) and multiplying on the right by γ^0 one finds

$$0 = \psi^\dagger(-i\overleftarrow{\partial}_\mu \gamma^{\mu\dagger} - m)\gamma^0 = -\psi^\dagger \gamma^0 (i\overleftarrow{\not{\partial}} + m)(\gamma^0)^2 = -\bar{\psi}(i\overleftarrow{\not{\partial}} + m)\,, \tag{3.18}$$

where $\bar{\psi} = \psi^\dagger \gamma^0$ denotes the *Dirac adjoint*. Then, for $J^\mu \equiv \bar{\psi}\gamma^\mu \psi$ one finds

$$\partial_\mu J^\mu = (\partial_\mu \bar{\psi})\gamma^\mu \psi + \bar{\psi}\gamma^\mu \partial_\mu \psi = -i\bar{\psi}(i\overleftarrow{\not{\partial}} + m)\psi - i\bar{\psi}(i\not{\partial} - m)\psi = 0\,, \tag{3.19}$$

so J^μ is a conserved four-vector current and as such a good probability four-current, with a manifestly positive zero component $J^0 = \bar{\psi}\gamma^0\psi = \psi^\dagger\psi$ which can be interpreted as a probability density.

Dirac Hamiltonian and energy eigenfunctions From Eq. (3.17) one obtains the Dirac Hamiltonian as follows,

$$i\partial_0 \psi = (m\gamma^0 - i\vec{\nabla}\cdot\gamma^0\vec{\gamma})\psi \equiv H_{\rm Dirac}\psi\,. \tag{3.20}$$

Plane wave solutions of the Dirac equations of the form $\psi = \psi_0 e^{-ip\cdot x}$ are easily seen to correspond to time-independent eigenfunctions $\psi_0 e^{i\vec{p}\cdot\vec{x}}$ of the Dirac Hamiltonian of energy p^0 and vice versa, since in both cases ψ_0 must satisfy

$$\begin{gathered}(i\not{\partial} - m)\psi_0 e^{-ip\cdot x} = e^{-ip\cdot x}(p^0\gamma^0 - \vec{p}\cdot\vec{\gamma} - m)\psi_0 \equiv e^{-ip\cdot x}(\not{p} - m)\psi_0 = 0\,,\\ \gamma^0(m + \vec{p}\cdot\vec{\gamma})\psi_0 = p^0\psi_0 \longrightarrow (p^0\gamma^0 - \vec{p}\cdot\vec{\gamma} - m)\psi_0 = (\not{p} - m)\psi_0 = 0\,.\end{gathered} \tag{3.21}$$

There are four solutions to Eq. (3.21), two with positive energy $p^0 = \sqrt{\vec{p}^2 + m^2}$ and two with negative energy $p^0 = -\sqrt{\vec{p}^2 + m^2}$. It is customary to fix $p^0 \equiv +\sqrt{\vec{p}^2 + m^2}$ and look for positive-energy $E = p^0$ solutions with momentum $\vec{p}$ of the form $\psi_+ = u(\vec{p})e^{-ip\cdot x}$, and for negative-energy solutions $E = -p^0$ with momentum $-\vec{p}$ of the form $\psi_- = v(\vec{p})e^{ip\cdot x}$, with u and v suitable bispinors satisfying the equations

$$(\not{p} - m)u(\vec{p}) = 0\,, \qquad (\not{p} + m)v(\vec{p}) = 0\,. \tag{3.22}$$

Setting

$$u = \begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}\,, \qquad v = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} \tag{3.23}$$

with two-component $\xi_{1,2}$ and $\eta_{1,2}$, we find

$$\begin{aligned}0 = (\not{p} - m)u &= \begin{pmatrix}(p^0 - m)\xi_1 - \vec{p}\cdot\vec{\sigma}\xi_2\\ -(p^0 + m)\xi_2 + \vec{p}\cdot\vec{\sigma}\xi_1\end{pmatrix}\,,\\ 0 = (\not{p} + m)v &= \begin{pmatrix}(p^0 + m)\eta_1 - \vec{p}\cdot\vec{\sigma}\eta_2\\ -(p^0 - m)\eta_2 + \vec{p}\cdot\vec{\sigma}\eta_1\end{pmatrix}\,.\end{aligned} \tag{3.24}$$

These are solved imposing that

$$\xi_2 = \frac{\vec{p}\cdot\vec{\sigma}}{p^0 + m}\xi_1\,, \qquad \eta_1 = \frac{\vec{p}\cdot\vec{\sigma}}{p^0 + m}\eta_2\,. \tag{3.25}$$

Choosing pairs of orthonormal spinors χ_s and $\tilde{\chi}_s$, $s = 1, 2$, $\chi_{s'}^\dagger \chi_s = \tilde{\chi}_{s'}^\dagger \tilde{\chi}_s = \delta_{s's}$, we can then write

$$u_s(\vec{p}) = \sqrt{p^0 + m} \begin{pmatrix} \chi_s \\ \frac{\vec{p}\cdot\vec{\sigma}}{p^0+m}\chi_s \end{pmatrix}, \qquad v_s(\vec{p}) = \sqrt{p^0 + m} \begin{pmatrix} \frac{\vec{p}\cdot\vec{\sigma}}{p^0+m}\tilde{\chi}_s \\ \tilde{\chi}_s \end{pmatrix}, \tag{3.26}$$

where the factor $\sqrt{p^0 + m}$ is chosen for normalisation purposes, so that

$$\begin{aligned} \bar{u}_{s'}(\vec{p})u_s(\vec{p}) &= 2m\delta_{s's}\,, & \bar{v}_{s'}(\vec{p})v_s(\vec{p}) &= -2m\delta_{s's}\,, \\ \bar{u}_{s'}(\vec{p})v_s(\vec{p}) &= 0\,, & \bar{v}_{s'}(\vec{p})u_s(\vec{p}) &= 0\,, \end{aligned} \tag{3.27}$$

where $\bar{u}_s(\vec{p}) = u_s(\vec{p})^\dagger\gamma^0$ and $\bar{v}_s(\vec{p}) = v_s(\vec{p})^\dagger\gamma^0$. Completeness of the solutions entails the relations

$$\sum_s u_s(\vec{p})\bar{u}_s(\vec{p}) = \not{p} + m\,, \qquad \sum_s v_s(\vec{p})\bar{v}_s(\vec{p}) = \not{p} - m\,. \tag{3.28}$$

Spin of the electron The double degeneracy of each energy level explains the two spin states of an electron. In fact, taking the low-energy limit $\vec{p} \to 0$ in the positive-energy solution we find

$$u_s(\vec{p}) \to \sqrt{2m} \begin{pmatrix} \chi_s \\ 0 \end{pmatrix}, \tag{3.29}$$

so only two components survive, which can be interpreted as the two components of the electron wave function. One usually takes

$$\chi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \chi_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{3.30}$$

so that the $s = 1$ solution corresponds to $s_z = \frac{1}{2}$ and the $s = 2$ solution corresponds to $s_z = -\frac{1}{2}$.

Antiparticles The need to interpret somehow the negative-energy solutions led to the prediction of the positron, i.e., the antielectron: a negative-energy solution of momentum $-\vec{p}$ for the electron is reintepreted as a positive-energy solution of momentum $\vec{p}$ for the positron. To see how this is possible, notice that we can write a negative-energy plane wave as follows,

$$e^{ip\cdot x} = e^{-i[p^0(-x^0)-(-\vec{p})\cdot\vec{x}]}\,, \tag{3.31}$$

which shows that we can reinterpret the negative-energy solution as a positive-energy solution travelling backwards in time. If we reverse the direction of time, consistency demands that we also flip the momentum and the spin of the particle, so a negative-energy solution of momentum $-\vec{p}$ and spin component $-s_z$ is reinterpreted as a positive-energy solution of momentum $\vec{p}$ and spin component s_z for a particle travelling backwards in time. In the low-energy limit we find for the negative-energy solution bispinor

$$v_s(\vec{p}) \to \sqrt{2m} \begin{pmatrix} 0 \\ \tilde{\chi}_s \end{pmatrix}, \tag{3.32}$$

and so we have to swap the roles of $\chi_{1,2}$ in Eq. (3.30) and set[34]

$$\tilde{\chi}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \tilde{\chi}_2 = -\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{3.33}$$

[34]The reason for the appearance of a minus sign is technical, and related to the usual conventions for the implementation of rotations on the states of a physical system. For more details see Section 4.2.4, in particular p. 94.

so that the $s = 1$ solution corresponds to $s_z = \frac{1}{2}$ and the $s = 2$ solution corresponds to $s_z = -\frac{1}{2}$.

A positive-energy particle travelling backwards in time can be further reinterpreted as a different type of particle travelling forwards in time, one that carries the exact opposite of every conserved charge (e.g., electric charge, baryon number,...) of the original particle. This is what we call an *antiparticle*: same mass and spin as the corresponding particle, but opposite charges. To make matter stable, one has to assume that the vacuum is actually a state where all the negative-energy states are occupied (*Dirac sea*), so preventing decays by Pauli's principle. An unoccupied negative-energy state gives then a net positive contribution to the energy of the system, and a net contribution to its charges opposite to that of a particle, so that it corresponds to a positive-energy antiparticle state.

This old-fashioned interpretation of negative-energy states, while working in practice, is quite unsatisfactory: what does it mean that a particle travels backwards in time? Where do the occupied negative-energy states come from? Moreover, the interpretation of negative-energy states as positive-energy states of an antiparticle would work also for the negative-energy solutions of the Klein-Gordon equation, but in that case one could not invoke Pauli's principle to ensure the stability of matter. We will see how to overcome these issues in the next Subsection, after completing the discussion of the solutions of the Dirac equation.

Chirality For future utility, notice that one can combine the four matrices in Eq. (3.14) into a fifth one,

$$\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} . \tag{3.34}$$

This anticommutes with all the others, $\{\gamma^5, \gamma^\mu\} = 0$. Since $(\gamma^5)^2 = \mathbf{1}$, one can decompose any solution of the Dirac equation into two components, $\psi_{R,L}$,

$$\psi = \frac{1+\gamma^5}{2}\psi + \frac{1-\gamma^5}{2}\psi = \psi_R + \psi_L \,, \tag{3.35}$$

with definite *chirality*,

$$\gamma^5\psi_R = \psi_R \,, \qquad \gamma^5\psi_L = -\psi_L \,, \tag{3.36}$$

called the right-handed and the left-handed components. These are exchanged under a parity transformation, i.e., an inversion of the spatial coordinates, and so should appear symmetrically in a parity-conserving theory.

General solution of the Dirac equation A momentum-space argument like the one we used for the Klein-Gordon equation shows that the plane wave solutions discussed above form a complete set of solutions of the Dirac equation. The most general solution is then obtained by taking a linear combination of these with generally complex coefficients, that reads

$$\psi(x) = \int d\Omega_p \sum_{s=1}^{2} \left\{ b_s(\vec{p}) u_s(\vec{p}) e^{-ip\cdot x} + d_s(\vec{p})^* v_s(\vec{p}) e^{ip\cdot x} \right\} , \tag{3.37}$$

where again $b_s(\vec{p})$ and $d_s(\vec{p})^*$ are the amplitudes of the normal modes, i.e., of the solutions with definite energy and spin component, corresponding respectively to positive- and negative-energy states, or to an electron and a positron according to our particle/antiparticle interpretation.

3.2 A sketch of Quantum Field Theory

As we have seen above, relativistic quantum mechanics has serious flaws: negative probabilities from the wave function of scalar particles, and negative energy states that we have to reinterpret as antiparticles in a not fully satisfactory way, invoking travelling backwards in time, and an infinite amount of occupied states to stabilise mattter (which moreover does not work for scalar particles). If one takes a step back (and perhaps with the benefit of hindsight) one sees that actually the whole premise of relativistic quantum mechanics is faulty. The wave function is in a sense a probability field - a field is something defined at all points in spacetime - telling us how likely it is to find a particle at a certain point at a certain time. In a relativistic process, though, particles can be created and annihilated: you make an electron and a positron collide and you obtain a proton and an antiproton - where have all the leptons gone? At some instant in time the electron and positron disappear *everywhere in space*, and from then on their wave function simply vanishes. This instant must be the same for the electron and positron if lepton number is to be conserved. But simultaneity is a frame-dependent property (unless the two events under consideration are also taking place at the same point in space - but the wave function must become zero everywhere), and in certain frames lepton number would not be conserved. Those frames would then be inequivalent, and Lorentz invariance would be lost. This should make clear that a description of relativistic processes based on one-particle wave functions will be inconsistent.

Another difficulty appears when introducing interactions. Finiteness of the speed of light tells us that interactions propagate at a finite speed and cannot act instantaneously at a finite distance. This prevents us from describing interactions via a potential in a relativistic setting. One way to comply with this is to make interactions local. For two pointlike objects, this means that they interact only if they are at the same place at the same time: this appears too limited. The next thing is to have the pointlike objects to interact locally with a field, which is defined everywhere in space and time, and to endow the field with dynamics that respect Lorentz invariance and the finiteness of the speed of light. In this way, it is the field that communicates the interaction from one object to another.

At this point, why not getting rid of the pointlike objects and describe everything with fields? After all, in quantum mechanics a "pointlike" particle is described by a wave function, and so it has an associated probability field. It is now easy to comply with the requirements of relativity if the equations of motion of the various fields are local, i.e., if the evolution of a field at a spacetime point x depends only on its own and other fields' value in the immediate vicinity of x – in other words, on fields and their derivatives at x. But where have the particles gone? These appear at a second stage, and correspond to localised excitations of these fields. Moreover, we can recover our "interaction by particle exchange" if we look at their interactions as "ripples" propagating between them through a suitable mediator field, which correspond to the exchange of a mediator particle. On the other, particles come in integer numbers only, unlike the amplitudes of field excitations: how can we deal with that?

Quantised fields The way out of these issues is to promote the probability field of the wave function to a different type of field, one which acts as an operator on the (Hilbert) space of states and can create or annihilate particles. This procedure is known as *second quantisation*. In practice (the theory is much more complicated) this amounts to upgrade the amplitudes of the field normal modes to operators that excite particles and add them to the state of the system,

starting out from the vacuum state (the state with no particles whatsoever); or erase them and remove them from the state.

For the case of a scalar particle, described by the Klein-Gordon equation, one has [see Eq. (3.8)]

$$\phi(x) = \int d\Omega_p \left\{ a(\vec{p}) e^{-ip\cdot x} + b(\vec{p})^\dagger e^{ip\cdot x} \right\} , \tag{3.38}$$

where now $a(\vec{p})$ and $b(\vec{p})^\dagger$ are operators, obeying the *commutation relations*

$$[a(\vec{p}), a(\vec{q})^\dagger] = (2\pi)^3 2p^0 \delta^{(3)}(\vec{p}-\vec{q}) , \qquad [b(\vec{p}), b(\vec{q})^\dagger] = (2\pi)^3 2p^0 \delta^{(3)}(\vec{p}-\vec{q}) , \tag{3.39}$$

all other commutators vanishing. These are the commutation relations of the annihilation and creation operators of two infinite sets of harmonic oscillators, with elements of each set labelled by the momentum $\vec{p}$. Exactly as a creation operator raises the energy level of a harmonic oscillator, $a(\vec{p})^\dagger$ and $b(\vec{p})^\dagger$ create respectively a particle and an antiparticle of momentum $\vec{p}$ out of the vacuum state; their Hermitian conjugate $a(\vec{p})$ and $b(\vec{p})$ destroy respectively a particle and an antiparticle. Applied to the state vector of the system, these operators increase or decrease the number of particles and antiparticles of momentum $\vec{p}$ by one unit. The Hamiltonian of the system is

$$H = \int d\Omega_p \, p^0 \left[a(\vec{p})^\dagger a(\vec{p}) + b(\vec{p})^\dagger b(\vec{p}) \right] , \tag{3.40}$$

which can be expressed in terms of the scalar field $\phi(x)$ and its spatial derivatives. In this context, the meaning of Q defined in Eq. (3.9) becomes clear. After quantisation one has

$$Q = \int d\Omega_p \left[a(\vec{p})^\dagger a(\vec{p}) - b(\vec{p})^\dagger b(\vec{p}) \right] , \tag{3.41}$$

and since $a(\vec{p})^\dagger a(\vec{p})$ and $b(\vec{p})^\dagger b(\vec{p})$ count respectively the number of particles and antiparticles of momentum $\vec{p}$, if one assigns electric charge 1 to particles and -1 to antiparticles then Q is the operator corresponding to electric charge. This should also make clear why $a(\vec{p})$ and $b(\vec{p})^\dagger$ appear in $\phi(x)$, instead of $a(\vec{p})$ and $b(\vec{p})$: since $a(\vec{p})$ annihilates a particle of charge 1 and $b(\vec{p})^\dagger$ creates a particle of charge -1, by applying $\phi(x)$ to a state of definite charge q one obtains a new state of definite charge $q-1$. This is made explicit by the commutation relation

$$[Q, \phi(x)] = -\phi(x) , \tag{3.42}$$

obeyed by the scalar field.

A similar procedure applies to the spin-$\frac{1}{2}$ particles described by the Dirac equation, leading to

$$\psi(x) = \int d\Omega_p \sum_{s=1}^{2} \left\{ b_s(\vec{p}) u_s(\vec{p}) e^{-ip\cdot x} + d_s(\vec{p})^\dagger v_s(\vec{p}) e^{ip\cdot x} \right\} , \tag{3.43}$$

with $b_s(\vec{p})$ and $d_s(\vec{p})^\dagger$ the annihilation and creation operators of the fermion and antifermion, respectively. For completeness, we report also the Dirac adjoint of this field,

$$\bar{\psi}(x) = \int d\Omega_p \sum_{s=1}^{2} \left\{ d_s(\vec{p}) \bar{v}_s(\vec{p}) e^{-ip\cdot x} + b_s(\vec{p})^\dagger \bar{u}_s(\vec{p}) e^{ip\cdot x} \right\} . \tag{3.44}$$

Analogously to the scalar case, one has for the Hamiltonian and the charge operator

$$\begin{aligned} H &= \int d\Omega_p \, p^0 \sum_{s=1}^{2} \left[b_s(\vec{p})^\dagger b_s(\vec{p}) + d_s(\vec{p})^\dagger d_s(\vec{p}) \right] , \\ Q &= \int d\Omega_p \sum_{s=1}^{2} \left[b_s(\vec{p})^\dagger b_s(\vec{p}) - d_s(\vec{p})^\dagger d_s(\vec{p}) \right] , \end{aligned} \tag{3.45}$$

which can be expressed in terms of ψ, $\bar{\psi}$, and their spatial derivatives. In this case, the fermionic nature of the system requires that the creation and annihilation operators obey *anticommutation relations*,

$$\{b_s(\vec{p}), b_{s'}(\vec{q})^\dagger\} = (2\pi)^3 2p^0 \delta_{ss'} \delta^{(3)}(\vec{p} - \vec{q}) , \qquad \{d_s(\vec{p}), d_{s'}(\vec{q})^\dagger\} = (2\pi)^3 2p^0 \delta_{ss'} \delta^{(3)}(\vec{p} - \vec{q}) , \tag{3.46}$$

all other anticommutators vanishing. In this way, particles created by $b_s(\vec{p})^\dagger$ and $d_s(\vec{p})^\dagger$ obey Pauli's principle, since $2\, b_s(\vec{p})^{\dagger\, 2} = \{b_s(\vec{p})^\dagger, b_s(\vec{p})^\dagger\} = 0$.[35] This approach can be extended also to the electromagnetic field, i.e., the photon field, which reads

$$A_\mu(x) = \int d\Omega_p \sum_{\lambda=1}^{2} \left\{ a_\lambda(\vec{p}) \epsilon_\mu^{(\lambda)}(\vec{p}) e^{-ip\cdot x} + a_\lambda(\vec{p})^\dagger \epsilon_\mu^{(\lambda)*}(\vec{p}) e^{ip\cdot x} \right\} , \tag{3.47}$$

where the four-vectors $\epsilon_\mu^{(1,2)}(\vec{p})$ correspond to the two physical polarisations of a photon, and $a_\lambda(\vec{p})$ and $a_\lambda(\vec{p})^\dagger$ again obey commutation relations, since the photon is a spin-1 boson. The same type of creation and annihilation operators appear in A_μ, corresponding to the fact that the photon is its own antiparticle.

The formalism discussed above adequately describes free particles. Incidentally, it also justifies why particles are indistinguishable: they all equally are excitations of the same field. Finally, it naturally leads (essentially forces us) to introduce antiparticles: if one is to assign some conserved charge to a quantum field, then its positive- and negative-energy components better be assigned the same charge. However, the positive-energy part annihilates a particle, removing some amount of charge from the state, and the negative-energy part creates a particle, adding some amount of charge to the state. In order to assign the same charge to the two components one then needs the creation operator appearing in the negative-energy part to create a particle with opposite charge to that associated with the other creation operator.[36] The quantum field language also dispenses us from the tricky intepretation of negative-energy states involving particles travelling backwards in time and the Dirac sea, treating particles and antiparticles in a more symmetric way.

Interactions While the formalism discussed above is adequate for free particles, what we really want is to describe interacting systems. A "simple" way to do so in this language is to include in the Hamiltonian certain combinations of (at least three) creation and annihilation

[35] As a matter of fact, the need to impose commutation relations when describing particles of integer spin, and anticommutation relations when describing particles of half-integer spin, originates in the requirements of Lorentz and translation invariance, and of positivity of the Hamiltonian. In other words, the bosonic or fermionic nature of particles and their connection with spin is a theorem in QFT (spin-statistics connection).

[36] This argument is made precise by the physical requirement of locality and the associated microcausality condition on quantum fields. For details, see Ref. [7].

operators that change the state and number of particles in the system, corresponding in practice to the emission or absorption processes mentioned in Section 1.3. This is done more easily by using fields rather than creation and annihilation operators directly, that, in a sense, create or annihilate (at least three) particles at the same point in spacetime. The advantage of using fields is that locality and symmetries of the interactions are manifestly under control, while achieving the same result using creation and annihilation operators is more cumbersome.

Consider for example the elastic interaction of two particles via electromagnetic forces. One can imagine that one of them emits a photon at some point (three subprocesses: the original particle is annihilated; a new particle of the same type and a photon are created), which is absorbed by the other (again three subprocesses: the original particle and the photon are annihilated; a new particle of the same type is created). As already mentioned in Section 1.3, each half of this process is called an interaction vertex. For an electron or a positron, in terms of the fermion field ψ, Eq. (3.43), its Dirac adjoint $\bar{\psi}$, and the photon field A_μ, the interaction Hamiltonian that allows us to describe this process reads

$$V_I(t) = e \int d^3x \, J^\mu(x) A_\mu(x) = e \int d^3x \, \bar{\psi}(x) \gamma^\mu \psi(x) A_\mu(x) \,, \tag{3.48}$$

that can expressed explicitly in terms of eight different combinations of creation and annihilation operators for the electron, positron, and photon. Among these there are two that realise precisely the elementary processes of emission and absorption of a photon by an electron. If we do not see the photon being exchanged between the electrons, the net effect is the interaction between the original two particles, whose state is changed by the electromagnetic force mediated by the photon. Notice that the use of fields and the constraints of locality and Lorentz invariance force us to introduce more elementary processes than the one we originally intended, corresponding to the other combinations of creation and annihilation operators appearing in the expansion of Eq. (3.48) that we mentioned above. We will discuss these in detail later in Section 3.3.

As a matter of fact, we had better *not* see the photon. In fact, energy and momentum cannot be conserved in a process where one particle comes in and two come out, or vice versa two come in and one comes out.[37] Violations of energy and/or momentum conservation are somehow acceptable if they take place on time and length scales smaller than what one can access according to the uncertainty principle, as we cannot experimentally observe such violations. If we insist on giving an interpretation in terms of real particles, we can invoke the uncertainty relation $\Delta x \Delta p \sim 1$, and translate it as follows as a relation between time and energy uncertainties:

$$1 \sim \Delta x \Delta p = \frac{\Delta x}{\Delta t} \frac{\Delta p}{\Delta E} \Delta E \Delta t = v \frac{1}{v} \Delta E \Delta t = \Delta E \Delta t \,, \tag{3.49}$$

where $\Delta x = v \Delta t$ relates the uncertainty Δt on when something happens to a particle to the position where it happens, assuming we know the speed v of the particle; $\Delta p / \Delta E = E/p$ follows instead from the relativistic energy-momentum relation $E = \sqrt{\vec{p}^2 + m^2}$. The relation above tells us that the exchanged photon can violate energy conservation by ΔE over a time $\Delta t \sim 1/\Delta E$, so that it can be emitted and absorbed. As we cannot resolve the process on a time scale shorter than what it takes for the (classically forbidden) emission and absorption processes to take place, we cannot say which electron emits and which one absorbs the photon.

[37] We could content ourselved with interactions where the same number of particles of each type come in and go out, but how could we describe creation and annihilation processes that way?

However, energy-momentum conservation is something we certainly do not want to abandon - at least not so easily,[38] not even for an intrinsically unobservable process. To reconcile it with our "particle-exchange picture" we have to posit that the exchanged photon is not on its mass shell $p_\gamma^2 = 0$: as it could not be observed anyway, this is not a big loss. In general, particles exchanged between other particles are only *virtual* particles, unobservable in any experiment, and not required to be on-shell – as opposed to *real* particle, that can be observed and must obey the mass-shell relation $p^2 = m^2$.

If we can now engineer the interactions, i.e., the combinations of fields mentioned above, in such a way that all the desired conservation laws are respected, then we have a candidate framework for the description of relativistic processes that avoids the conceptual difficulties of relativistic quantum mechanics (if one accepts virtual particles - but we will shortly see that one does not even have to do that). The way to enforce conservation laws is through their close connection with symmetry: the symmetries of a system in fact imply the existence of conserved quantities – build interactions that respect the desired symmetries, and you have your conservation laws. The key point is to put the fields at the centre of the description, and obtain particles as excitations of the fields upon their quantisation. As we already remarked above, working with fields one can build interactions with the desired locality and symmetry properties much more easily than working directly with creation and annihilation operators.

In this framework, the picture of "interaction through particle exchange" depicted above emerges naturally when treating the interaction Hamiltonian as a perturbation added to the free Hamiltonian, describing non-interacting particles. Terms in the perturbative expansion of the S-matrix [see comment after Eq. (2.60) on page 35] are built by joining two or more basic interaction vertices, with the particles defining the particular process one is looking at being the real particles that enter and exit the scene; and the particles being exchanged being the virtual particles that never make it to the detectors. Virtual particles are so virtual that are in effect just a computational device in a specific approximation scheme - but nonetheless allow for an intuitively clear picture of particle interactions, one to which we will stick from now on.[39]

The derivation of the particle-exchange picture in quantum field theory gives it also a quantitative meaning. One can in fact associate a complex number with each and all of the infinitely many possible ways that a specific physical process (one with prescribed sets of particles in the initial and final states) can take place. For the elastic process mentioned above, exchange of one, two, n photons; which in turn may split into a particle-antiparticle pair, that goes on with its business until eventually annihilating into a photon again; and so on. Each of these ways is called a *Feynman diagram*. The sum of all these complex numbers provides the scattering amplitude, from which the scattering probability and eventually the scattering cross section are obtained. Computing the complex number associated with a specific Feynman diagram is made easy by a set of rules (the *Feynman rules*) that translate each part of the diagram in a well defined mathematical quantity, and tell us how to combine them together. A proper derivation of these rules requires first a proper development of quantum field theory, which is outside of

[38] *Any* fundamental assumption may have to be abandoned at some point, and survive only as an approximate assumption valid in certain length/energy regimes.

[39] One may wonder what happens if the perturbative approach is not allowed by the strength of the interaction Hamiltonian, which may radically change the scene compared to the free one. This is the case for strong interactions, where the free quark Hamiltonian turns into an interacting Hamiltonian that asymptotically describes free *hadrons*. In this case one cannot generally use perturbation theory to describe interactions as small modifications of the state of affairs. Surprisingly, one is allowed to use use perturbation theory to describe strong interactions very accurately at high energy.

the scope of this course. We will be content with listing them, and using them for a quantitative or semi-quantitative understanding of particle physics.[40]

3.3 Feynman diagrams

We now discuss the program outlined above in some detail. The way this works in practice is that one computes the matrix elements of the S operator between initial and final multiparticle states by expanding Dyson's formula for the S-matrix, Eq. (2.60), and computing term by term the various contributions, i.e.,

$$\langle f|S-\mathbf{1}|i\rangle = (-i)\int dt\,\langle f|V_I(t)|i\rangle + \frac{(-i)^2}{2}\int dt_1\int dt_2\,\theta(t_1-t_2)\langle f|V_I(t_1)V_I(t_2)|i\rangle + \dots\,. \quad (3.50)$$

The interacting Hamiltonian in the interaction picture is built out of the same creation and annihilation operators that create and destroy free particles out of the vacuum. For each term in Eq. (3.50) one has to evaluate the matrix element of a linear combination of products of creation and annihilation operators between particle states, which can then be obtained explicitly. In doing so, one needs to take into account that such operators can affect any of the particles of the corresponding type present in the initial and final states.

This calculation can be efficiently carried out with the aid of a graphical device. With each factor of $V_I(t)$ in Eq. (3.50) we associate a graph with as many lines as fields (and so creation and annihilation operators) in the interacting Hamiltonian, with a different type of line for each type of field (and so type of particle). All these lines meet at a point, which we also call interaction vertex. As an example, in Fig. 8 (left) we show the electromagnetic interaction vertex involving two fermionic fields and one photon field, describing the creation/annihilation of a fermion or an antifermion and a photon. The fermion field is associated with a solid line, and the photon field with a wavy line. The term of order n in Eq. (3.50) then contains n interaction vertices. The lines attached to the various vertices are then paired with the incoming and outgoing particles of the same type in all possible ways, resulting in an *external line* for the graph; the remaining lines of each type are further paired with each other in all possible ways, resulting in an *internal line* for the graph. If there are still initial and final particles not paired with a line, we pair them with each other compatibly with their type, and add a line running across the graph, which can be thought of as a pair of external lines. This corresponds to one or more of the initial particles going through the scattering process not interacting with the rest and remaining unaffected.

The meaning of these pairings is intuitively clear: each time we join a line with an incoming or outgoing particle or with another line, we are essentially pairing one creation and one annihilation operator among those appearing in the fields and in the initial and final state vectors, so describing the propagation of the corresponding particle from one point of the graph to another. Each combination of pairings corresponds then to a process in which virtual particles are exchanged in an arbitrarily complex way between the incoming and outgoing particles. As an example, in Fig. 8 (right) we show a possible virtual process with two interaction vertices for the elastic scattering of two electrons, involving the exchange of a virtual photon. It is mandatory that no line or particle remain unpaired: this would mean that there is a mismatch between the particles in the initial state, those in the final state, and the net amount of particles created or annihilated by the interaction, meaning that no virtual exchange process with the given amount of interaction vertices exists for the physical scattering process under consideration.

[40] Paraphrasing J. Schwinger, Feynman rules brought quantum field theory to the masses.

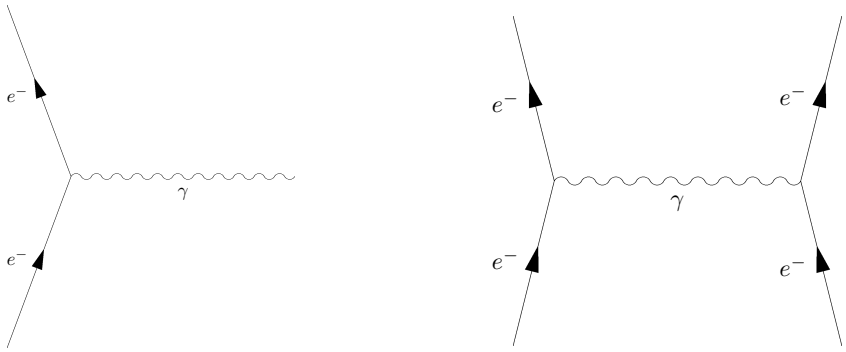

Figure 8: Left: the electromagnetic interaction vertex. Right: exchange of a photon between two electrons.

Each of the resulting distinct graphs is a Feynman diagram. To each of its elements is associated a mathematical counterpart according to well defined rules, which are of the following schematic form:

- each external line contributes a *wave function* (this is only a name, not to be confused with the particle wave function in quantum mechanics), depending on the type and on the quantum numbers (momentum, spin component) of the corresponding real (on-shell) particle;
- each internal line, representing a virtual (off-shell) particle, contributes what is called a *propagator*, again depending on the type of particles, and carrying similar quantum numbers; in general, a propagator contains a factor $\frac{1}{p^2-m^2}$, with p the four-momentum carried by the virtual particle and m its (true) mass;
- each vertex contributes a *coupling constant* factor, and typically some matrix-type object that combines together wave functions and propagators, determining how quantum numbers are carried from one external line to another by the internal lines – i.e., how quantum numbers are exchanged through the virtual particles;
- in particular, conserved quantities (E, $\vec{p}$, angular momentum, electric charge,...) are conserved at each vertex, and so everywhere – ensuring their conservation from the initial to the final state.

Further, more technical rules are also needed, but do not change the general picture.

Let us return to the electromagnetic vertex of Fig. 8 (left). This describes the emission or absorption of a photon, corresponding to the wavy line, by an electron, corresponding to the oriented solid line. By convention, we choose time to flow upwards, so that the orientation of the electron line is the same as that of the time arrow. The addition of an orientation to the fermion line is done to distinguish between electrons and positrons, which are described by the same field but are different particles. In Eq. (3.48), there are four possible combinations of fermionic annihilation and creation operators, that we can write schematically as $b^\dagger b$, $dd^\dagger$, db, and $b^\dagger d^\dagger$. The first combination is the one represented in Fig. 8 (left), reproduced also in Fig. 8 (top left), with an electron first annihilated and subsequently created. The other combinations are depicted in Fig. 9, by representing a positron by a line oriented in the direction opposite

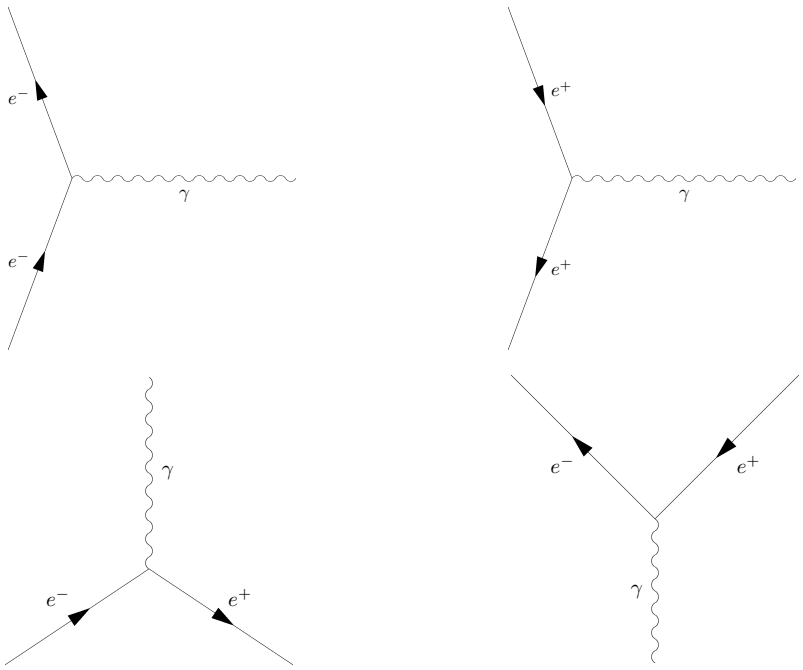

Figure 9: The interaction vertex of QED in its various guises.

to that of time: annihilation and subsequent creation of a positron (top right), annihilation of an electron and a positron (bottom left), and creation of an electron and a positron (bottom right). When joining lines from different vertices, only those parts containing the same type of creation/annihilation operator can be paired, and so a definite orientation can be associated with each fermion line. This is intuitively clear, as each line in a Feynman diagram corresponds essentially to the propagation of particle (either real of virtual) of a definite type, annihilated or created at some point, and created or annihilated at another point. Since the photon is its own antiparticle, an orientation is not needed for a photon line.

Let us return now to the photon exchange diagram of Fig. 8 (right). The relevant combination of fermionic operators $b^\dagger b$ in each vertex appears multiplied by either a or $a^\dagger$, i.e., the annihilation and creation operators of the photon. The exchange of a photon means that this is created at one vertex and annihilated at the other - but what happens at which vertex? Is there a difference between the two possibilities? A detailed analysis shows that both equally contribute to the scattering amplitude. This fits in with our qualitative considerations about the particle-exchange picture: since the photon is virtual, we do not and cannot know what happens at what vertex, and so we have to consider both possibilities on the same footing.

3.4 Interaction vertices of the fundamental interactions

Before discussing how to combine them into Feynman diagrams representing physical processes, we present the various interaction vertices characterising the various fundamental interactions. We will freely say that particles are emitted/absorbed at a vertex, as the problems of the classical interpretation of such a process are of no concern, as already discussed in detail.

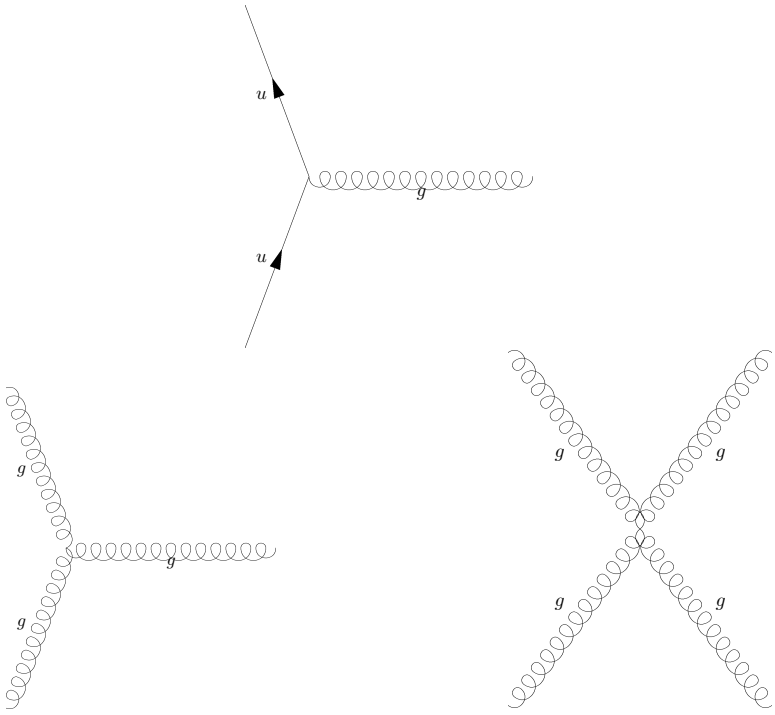

Figure 10: Interaction vertices of QCD.

Electromagnetic interactions There is only one interaction vertex in QED, that of Fig. 8 (left), up to rotations and reflections. This is reproduced in Fig. 9 (top left), and describes emission/absorption of a photon by an electron. Flipping it along the time direction, one describes emission/absorption of a photon by an antielectron (positron), Fig. 9 (top right). Rotating it with the photon line up, it describes electron-positron annihilation into a photon, Fig. 9 (bottom left); flipping it so that the photon line is down, it describes electron-positron pair creation from a photon, Fig. 9 (bottom right). As already discussed at length, these are not physically admissible processes if considered in isolation. For each vertex there is a factor e in the corresponding contribution to the transition amplitude. Energy, momentum, angular momentum, electric charge, as well as the particle number (number of particles minus antiparticles) of each type are conserved at a vertex. All electrically charged particles interact through this type of vertex, requiring only the modification $e \to Q$ of the coupling constant, where Q is the electric charge of the type of particle involved.

Strong interactions In QCD there is a vertex similar to the QED one, with a quark emitting or absorbing a gluon, without changing its type, or *flavour*, see Fig. 10 (top). The main difference is that quarks and gluons carry an extra degree of freedom, called *colour*. The emission/absorption of a gluon can change the colour of the quark, but this quantity is overall conserved at the vertex, together with energy, momentum, angular momentum, electric charge, and flavour. Quarks come in three colours, and there are eight types of gluons corresponding to the various colour-changing possibilities (of the nine possible colour combinations, the one corresponding to gluons that leave all the colours unchanged does not appear). Each such vertex has an associated factor g_s – the strong coupling constant. Similarly to the electromagnetic

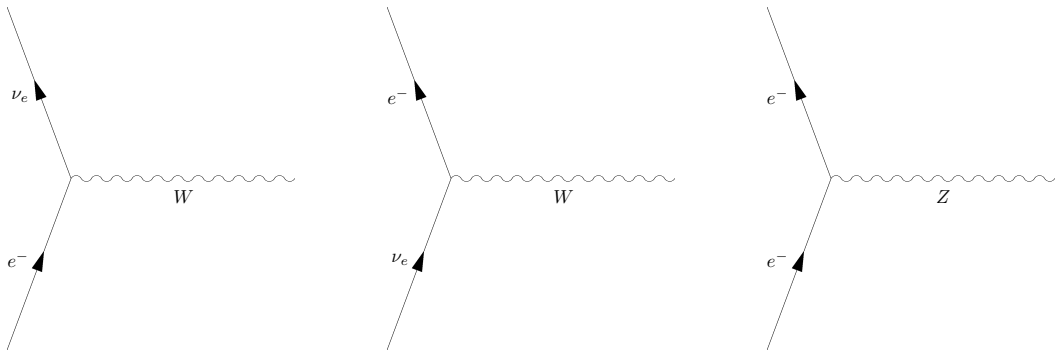

Figure 11: Interaction vertices of the weak interactions: charged (left and centre) and neutral current (right).

case, the interaction vertex can be flipped upside down to describe an antiquark emitting or absorbing a gluon, or turned around to describe quark-antiquark annihilation into a gluon or quark-antiquark pair production from a gluon. Differently from photons, gluons self-interact through the three- and four-gluon vertices of Fig. 10 (bottom), with associated factors g_s and g_s^2, respectively: this makes QCD a very different theory from QED. Only quarks, antiquarks and gluons are affected by strong interactions.

It is worth mentioning that, since quarks and gluons are permanently bound into hadrons due to confinement, what is effectively exchanged to mediate the strong interactions are hadrons instead of individual quarks and gluons. In particular, the lightest mediator is therefore not the massless gluon, but the massive (although relatively light) pion. This affects certain macroscopic properties of the interaction (see p. 11 in Section 1.3, and Section 3.5 below).

Weak interactions Weak interaction vertices involving leptons are of two types: the *charged current* vertex, with a negatively charged lepton $\ell^- = e^-, \mu^-, \tau^-$ emitting a W^--boson (or absorbing a W^+-boson) and turning into a neutrino of the same family, $\nu_\ell = \nu_e, \nu_\mu, \nu_\tau$, Fig. 11 (left); or a neutrino absorbing a W^--boson (or emitting a W^+-boson) and turning into the corresponding negatively charged lepton, Fig. 11 (centre); and the *neutral current* vertex, with the lepton emitting a Z boson, Fig. 11 (right). The vertically flipped version of the charged-current vertex in Fig. 11 (left) describes a positively charged lepton $\ell^+ = e^+, \mu^+, \tau^+$ turning into the corresponding antineutrino $\bar{\nu}_\ell = \bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$, through emission/absorption of a $W^{+/-}$-boson; similarly, the flipped version of Fig. 11 (centre) describes and antineutrino turning into the corresponding positively charged lepton. The various rotated versions of these vertices describe annihilation and pair production of $\ell^- \bar{\nu}_\ell$ or $\ell^+ \nu_\ell$ pairs. Flipping and rotating the neutral-current vertex yields results entirely analogous to the electromagnetic case. The two couplings g_w and g_w^0, associated respectively with the charged and neutral interaction vertices, are actually proportional to the electromagnetic coupling e via two functions of one parameter, the weak (or Weinberg) angle θ_W,[41] i.e., $g_w = e/\sin\theta_W$ and $g_w^0 = g/\cos\theta_W = 2e/\sin 2\theta_W$. While energy, momentum, angular momentum, and electric charge are conserved, the particle type is conserved by neutral-current interactions but is changed in charged-current interactions, and only the *lepton family number* (i.e., number of electrons/muons/tau and corresponding

[41] Introduced by Glashow.

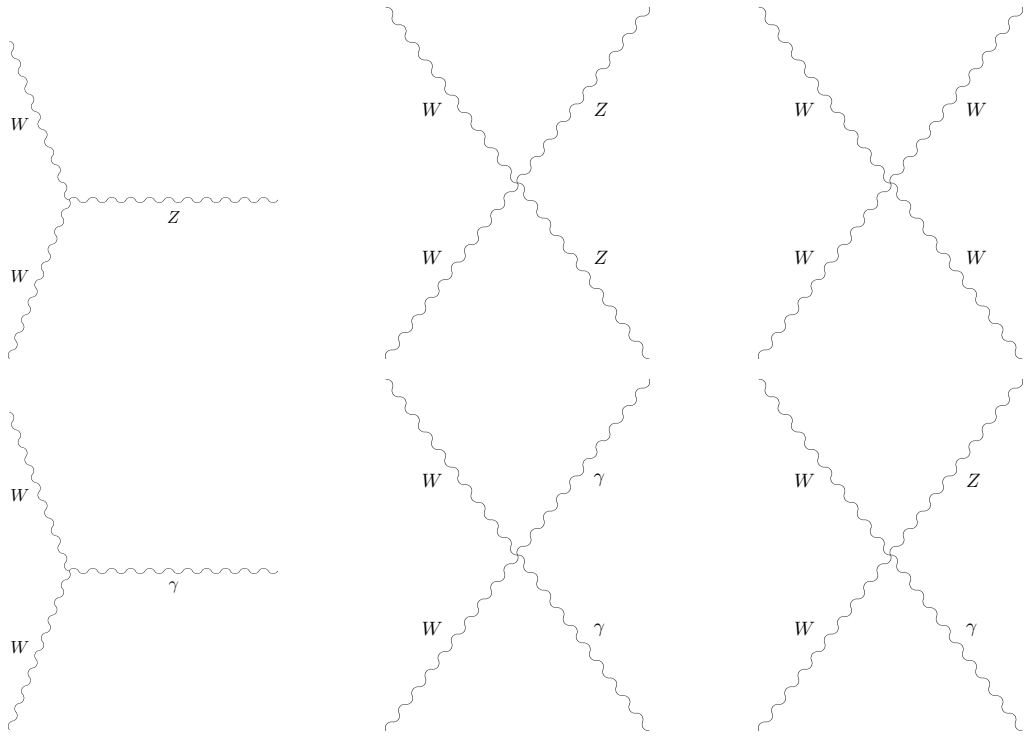

Figure 12: Interaction vertices of the weak interactions involving only intermediate vector bosons and photons.

neutrinos minus number of antielectrons/antimuons/antitau and corresponding antineutrinos) remains conserved.

The same two types of vertices described above also affect quarks, but while the weak neutral current works exactly in the same way as with leptons, there is a twist in the case of the charged current. The exact analogue of the leptonic process involving the weak charged current would be that only quarks in the same family are coupled: for example, a u quark emitting a W^+ and turning into a d. This interaction would change the flavour of quarks but only within a family, and it would therefore be possible to define a conserved "quark family number" in analogy with what we discussed above for leptons. It turns out, however, that after emitting a W^+ the up quark does not turn simply into a down quark, but into a linear combination of down, strange and bottom, $d' = \alpha d + \beta s + \gamma b$. In other words, the quark strong (mass) eigenstates do not coincide with their weak eigenstates. If this did not happen, it would be impossible to explain the hadronic decays of the kaons, where a strange (anti)quark turns into an up (anti)quark. The unitary matrix that determines how flavours mix is called the *Cabibbo-Kobayashi-Maskawa (CKM) matrix*.

Finally, there are also vertices not involving matter particles, but only intermediate vector bosons and photons (Fig. 12). This means that every elementary particle is affected by weak interactions, except for gluons.

3.5 Range and strength of the interactions

The general idea of associating fields with the elementary particles and the schematic description of Feynman rules given above already suffice to put on more solid ground our qualitative discussion of the range and strength of the interactions in Section 1.3.

Range of interactions (reprise) Consider the very non-relativistic, static (time-independent) limit of motionless matter interacting electromagnetically. Differently from the case of high-energy particles, passing fast next to each other and having the chance to exchange just a few photons, static charges can exchange a very large number of photons. In this limit the quantised nature of the photon field gets blurred and becomes practically undetectable, so that this field reduces to the familiar, classical electromagnetic field. In this case the interaction between electrically charged static particles is described in terms of the electric potential $V(\vec{x}\,)$. The potential generated by a static particle of charge Q obeys the well-known equation

$$-\Delta V(\vec{x}\,) = Q\delta(\vec{x}\,)\,, \tag{3.51}$$

solved by the familiar Coulomb potential $V = \frac{Q}{4\pi r}$, describing the interaction between a charged probe and our static charge. This equation is the static limit of the relativistic Maxwell equation in Coulomb gauge,

$$\Box V(t,\vec{x}\,) = \rho(t,\vec{x})\,, \tag{3.52}$$

valid for a time-independent and point-like charge distribution ρ, for which a time-independent potential provides the solution.

Play now the same game starting from the relativistic Klein-Gordon equation for a massive scalar field, known also as *Yukawa field*, describing particles of mass M. In the same static limit described above, the Yukawa field becomes a classical field describing the instantaneous interaction of matter particles that are coupled to it. In the presence of a static source of charge g, the static limit of the Klein-Gordon equation is

$$(-\Delta + M^2)u(\vec{x}\,) = g\delta(\vec{x}\,)\,, \tag{3.53}$$

which is the analogue of Eq. (3.51) but for a massive interaction field. The solution $u(\vec{x}\,)$ provides the potential associated with a massive scalar field in the non-relativistic limit. The solution of Eq. (3.53) is obtained most easily after making a Fourier transform to momentum space,

$$u(\vec{x}) = \int \frac{d^3p}{(2\pi)^3}\, e^{i\vec{p}\cdot\vec{x}}\tilde{u}(\vec{p}\,)\,, \tag{3.54}$$

where the equation reads

$$(\vec{p}^{\,2} + M^2)\tilde{u}(\vec{p}\,) = g\,. \tag{3.55}$$

This is an algebraic equation that is easily solved,

$$\tilde{u}(\vec{p}\,) = \frac{g}{\vec{p}^{\,2} + M^2}\,. \tag{3.56}$$

Going now back to coordinate space we get ($r = |\vec{x}|$)

$$\begin{aligned} u(\vec{x}) &= \int \frac{d^3p}{(2\pi)^3}\, e^{i\vec{p}\cdot\vec{x}} \frac{g}{\vec{p}^2 + M^2} = \frac{g}{(2\pi)^2} \int_0^\infty dp\, p^2 \int_{-1}^{+1} dz\, e^{iprz} \frac{1}{p^2 + M^2} \\ &= \frac{g}{(2\pi)^2 ir} \int_0^\infty dp\, p\, (e^{ipr} - e^{-ipr}) \frac{1}{p^2 + M^2} = \frac{g}{(2\pi)^2 ir} \int_{-\infty}^\infty dp\, p\, e^{ipr} \frac{1}{p^2 + M^2} \\ &= \frac{g}{(2\pi)^2 ir} (2\pi i) e^{-Mr} \frac{iM}{2iM} = \frac{g}{4\pi r}\, e^{-Mr}\,, \end{aligned} \tag{3.57}$$

where we have made use of the residue theorem to compute the last integral. The resulting central potential $u(\vec{x}) = \frac{g}{4\pi r}\, e^{-Mr} \equiv V_{\text{Yukawa}}(r)$ is the *Yukawa potential*, and it clearly has range $1/M$. This shows that the inverse of the mediator mass is indeed determining the range of the corresponding interaction. Putting back the appropriate factors of $\hbar$ and c we get Eq. (1.1), and substituting the W and pion masses we get Eqs. (1.2) and (1.3).

The factor corresponding to an internal line in a Feynman diagram is in a sense the relativistic analogue of the interaction potential discussed above, as it represents the interaction as the exchange of a particle. It is not by chance then that this is proportional to $(p^2 - m^2)^{-1} = (p_0^2 - \vec{p}^2 - m^2)^{-1}$, with p the momentum flowing through the internal line, which up to factors reduces to (3.56) in limit $p_0 \to 0$ in which the exchanged particle has very low energy (or, correspondingly, very long wavelength). In fact, one could compute the relativistic scattering amplitude corresponding to the exchange of one scalar particle, where the propagator appears, take the low-energy limit, and compare it with the lowest-order (Born) approximation for non-relativistic quantum-mechanical scattering, where the interaction potential appears. Matching the two one finds that the resulting non-relativistic potential is precisely the Yukawa potential.

Strength of interactions (reprise) We now turn to the strength of the various interactions. A simple way to estimate the relative strength of forces is to compare the lifetimes of unstable particles whose decayes are governed by different interactions. As we have said above, each Feynman diagram related to a certain physical process gives a contribution to the corresponding transition amplitude, whose absolute value square transition probability of the process. This in turn is proportional to the cross section of a scattering process, or to the decay width of a decay process.[42] In particular, the lifetime $\tau = \Gamma^{-1}$ of an unstable particle is then obtained qualitatively as $\tau \propto |\text{amplitude}|^{-2}$. The simplest diagrams for decay processes typically involve two vertices: for example, a muon emits a W^- turning into a muonic neutrino through the muonic analogue of the vertex in Fig. 11 (left), and the W-boson subsequently produces an electron-electronic antineutrino pair through the "rotated" version of this vertex. When the coupling is small, these diagrams give the leading contribution to the decay width, and so Γ is typically proportional to the fourth power of the coupling constant. One has then

$$\frac{\tau_2}{\tau_1} \sim \frac{g_1^4}{g_2^4} \quad \Longrightarrow \quad \frac{g_1^2}{g_2^2} \sim \sqrt{\frac{\tau_2}{\tau_1}}\,, \tag{3.58}$$

which allows us to compare the strength of the various interactions by comparing the typical lifetimes of the decay processes they cause.

[42] Recall that for fixed-target experiments the probability that a scattering event takes place is determined empirically as $N_{\text{events}}/(N_b N_t) = \sigma/A_b$. For a decay process, the probability of a decay is obtained empirically as $N_{\text{events}}/N_{\text{particles}} = \Gamma T$ where $N_{\text{particles}}$ is the sample size and T the observation time.

As we have already said, typical decay times for processes governed by strong interactions are in the range $10^{-23} \div 10^{-20} s$, while for weak interactions they are in the range $10^{-13} \div 10^{3} s$, and for electromagnetic interactions they are around $10^{-16} s$. This already tells us the ranking in strength of the various interactions. For a more detailed comparison, for strong and weak interactions we can for example compare the decay of the Δ^0, which takes place through strong interactions (mainly $\Delta^0 \to p\,\pi^-, n\,\pi^0$) with $\tau_\Delta = 5.6 \cdot 10^{-24} s$, to that of the neutron, which takes place through weak interactions (mainly $n \to p\,e^-\,\bar{\nu}_e$) with $\tau_n = 880\,s$:

$$\frac{g_w^2}{g_s^2} \sim \sqrt{\frac{\tau_\Delta}{\tau_n}} \sim 10^{-13}\,. \tag{3.59}$$

To compare electromagnetic and weak interactions we can for example compare the decay of the neutral pion, which is governed by electromagnetic interactions (mainly $\pi^0 \to \gamma\gamma$) with $\tau_{\pi^0} = 8.4 \cdot 10^{-17} s$, with the decay of the charged pion, which is governed by the weak interactions (mainly $\pi^+ \to \mu^+ \nu_\mu$) with $\tau_{\pi^+} = 2.6 \cdot 10^{-8} s$. We find

$$\frac{g_w^2}{g_{em}^2} \sim \sqrt{\frac{\tau_{\pi^0}}{\tau_{\pi^+}}} \sim 10^{-4} \div 10^{-3}\,. \tag{3.60}$$

The estimates Eqs. (3.59) and (3.60) will change if one uses different pairs of processes, but the bottom line is that the strong force is stronger than the electromagnetic one which is stronger than the weak one, with a clear hierarchy.

A similar kind of estimate can be done looking at scattering processes. The cross section σ of a process is in fact proportional to the absolute value squared of the amplitude, and computing this using the simplest Feynman diagrams one typically finds that σ is proportional to the fourth power of the coupling constant. We then find that the stronger the interaction, the larger the typical cross section for processes mediated by it, as we had anticipated.

It is worth mentioning at this point that the reason why weak interactions are much weaker than the electromagnetic ones is not really due to a weaker coupling constant, and that the estimate given above is a bit misleading. In fact, in the Glashow-Weinberg-Salam theory that describes weak and electromagnetic interactions in a unified fashion, the weak and electromagnetic coupling constants g_w and $g_{em} = e$ are related, and actually $e^2 \sim 0.2 g_w^2$! The main difference between the two interactions is that while the photon is massless, the intermediate vector bosons are very massive. As we have mentioned above, the factor corresponding to an internal line in a Feynman diagram is proportional to $(p^2 - m^2)^{-1}$, with p and m the four-momentum and the mass of the virtual particle, respectively. For processes at energies much lower than m, one can neglect p^2, and the propagator reduces to a p-independent factor. Effectively, the exchange of a W boson between a pair of particles with energies much lower than the W mass comes with a factor $g_w^2/M_W^2 \propto G_F =\sim 1.1 \cdot 10^{-5}$ GeV^{-2}, known as *Fermi constant*: this is the coupling constant actually estimated above. More precisely, the dimensionless quantity "g_w^2" appearing in Eqs. (3.59) and (3.60) is effectively G_F times some mass-squared scale M^2, that depends on the details of the process and can vary over a large range. This explains the very wide range of lifetimes found for unstable particles decaying through weak interactions. Weak interactions at low energies are therefore weak not because of a small coupling constant g_w, but because of the large mass of the mediators that lead to a small effective coupling "g_w^2" $= g_w^2 (M/M_W)^2$. This applies as long as we work at energies well below the $W^\pm$ mass, when the p^2 term can be neglected compared to the mass in the internal boson lines. When the two terms are comparable the weak interaction is actually much stronger than the electromagnetic one.

3.6 Basic physical processes and Feynman rules

The translation of Feynman diagrams into expressions contributing to S-matrix elements follows from the application of a set of rules, that can be derived in detail within the formalism of QFT. Here we describe them with the aid of the simplest physical processes.

The diagrams of Figs. 9 to 12 describe how the interaction works at the most fundamental level. However, matter particles appear only in three-particle vertices which, as discussed above, cannot represent true physical processes due to energy-momentum conservation: to describe an actual physical process one has to properly combine two or more of them.

3.6.1 Electromagnetic interactions

Let us begin with electromagnetic interactions. As we already saw above, if we combine two electromagnetic vertices together like in the left panel of Fig. 8, then we are representing the scattering process of two electrons (*Møller scattering*). This is reproduced in the left panel of Fig. 13, together with another way to realise the same process in the right panel: since electrons are indistinguishable, we cannot possibly know which one is coming out of the process going (say) left, and so we have to take into account both possibilities. These are all the diagrams with two interaction vertices. Each contains four external electron legs, and one internal photon line. The electron lines appear in two pairs forming an uninterrupted oriented line connecting the initial and final state of the process.

As we explained above, each Feynman diagram corresponds to a way of choosing creation and annihilation operators out of the field operators appearing in the interaction Hamiltonian in the interaction picture, Eq. (3.48). From the explicit expression of the fermion fields ψ and $\bar{\psi}$, Eqs. (3.43) and (3.44), we see that there is only one annihilation and one creation operator both for electrons and positron at each vertex. This means that there is only one way to pair a line coming out of an electromagnetic vertex with an initial electron: one has to choose the line corresponding to ψ, so that it contains the appropriate annihilation operator. Similarly, there is only one way to choose the line so that it contains the appropriate annihilation or creation operator to take care of the other possibilities: for an electron in the final state we need the line corresponding to $\bar{\psi}$; for a positron in the initial state we need again the line corresponding to $\bar{\psi}$; and for a positron in the final state we need the line corresponding to ψ. The counting of the possible equivalent ways to pair vertices and external lines is then straightforward. It should be clear also that if we have n interaction vertices, once we have made a pairing of lines and particles, one can always relabel the vertices and obtain the exact same contribution. This yields a factor $n!$ that compensates the $1/n!$ appearing in Dyson's formula.

Pairing lines with initial and final particles means essentially that we choose which annhihilation or creation operator does the job of removing a particle from the initial or final state. This leaves behind the scalar product $\langle 0|0\rangle = 1$, and the product of the coefficient multiplying said operators. Looking again at Eqs. (3.43) and (3.44), we see that when a vertex at x is paired with an initial particle of momentum p one gets a factor $e^{-ip\cdot x}$, while when paired with a final particle of momentum p' one gets a factor $e^{ip'\cdot x}$. Since we are integrating both over time [see Eq. (2.60)] and space [see Eq. (3.48)], the net effect is

$$\int d^4x\, e^{i\left(\sum_{\text{final}} p' - \sum_{\text{initial}} p\right)\cdot x} = (2\pi)^4 \delta^{(4)}\left(\sum_{\text{final}} p' - \sum_{\text{initial}} p\right), \tag{3.61}$$

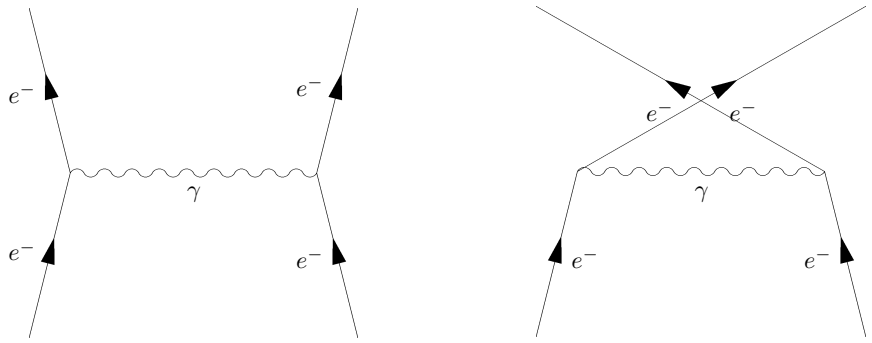

Figure 13: Møller scattering.

where the sums are over the final and initial particles directly connected to the vertex at x by an external line. What we get is then momentum conservation at each vertex. Moreover, these phase factors are accompanied by the appropriate solution of the Dirac equation, Eq. (3.26). Finally, since fields are multiplied at a vertex according to Eq. (3.48), one has that these solutions will appear multiplied on the left or on the right by a factor γ^μ.

To obtain the corresponding contribution to the amplitude one then proceeds as follows.

- Consider first the left diagram. Start from the end of a fermionic line. In this case these correspond to the electron legs in the final state. Start from, say, the leftmost one, and write down the wave function $\bar{u}_{s_1'}(\vec{p}_1')$ corresponding to the quantum numbers of the final electron on the left.

- Follow the line backwards. When encountering a vertex, include a vertex factor

$$-ie\gamma^\mu\,.$$

  The coupling constant e determines the strength of the interaction. The gamma matrix index μ correspond to the polarisation of the exchanged (virtual) photon.

- Continue following the fermion line backward. In this case, one immediately encounters another external leg, corresponding to the intial state of electron 1. Include then the corresponding wave function $u_{s_1}(\vec{p}_1)$. The result so far is

$$\bar{u}_{s_1'}(\vec{p}_1')(-ie)\gamma^\mu u_{s_1}(\vec{p}_1)\,,$$

  where all the matrix indices are contracted in the usual way.

- Move on to the end of another fermion line, which in this case is the other electron leg in the final state. Again, write down the wave function $\bar{u}_{s_2'}(\vec{p}_2')$ corresponding to the state of the second electron in the final state, then follow back until the second interaction vertex, and include a vertex factor

$$-ie\gamma^\nu\,,$$

  where the index ν correspond to the (possibly different) polarisation of the (virtual) photon after the exchange. As before, proceed along the fermion line, encountering the second

initial electron leg, and include the corresponding wave function $u_{s_2}(\vec{p}_2)$. The resulting factor is

$$\bar{u}_{s_2'}(\vec{p}_2')(-ie)\gamma^\nu u_{s_2}(\vec{p}_2)\,,$$

that multiplies the one found above.

- The last step is to include the factor corresponding to the internal photon line, i.e., the photon propagator. This reads

$$D_{\mu\nu}(q) = -i\frac{\eta_{\mu\nu}}{q^2 + i\epsilon}\,.$$

Here q is the momentum carried by the virtual photon (so that $q^2 \neq 0$ in general). Conservation of four-momentum at each vertex implies $q = p_1' - p_1 = p_2 - p_2'$ (the direction of the momentum, and so the sign of q, is irrelevant, as it should be). The quantity ϵ is an infinitesimal positive number which will be set to zero at the end of the calculation.

- Putting everything together, one finds

$$\frac{ie^2}{q^2 + i\epsilon}[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_2'}(\vec{p}_2')\gamma_\mu u_{s_2}(\vec{p}_2)]\,.$$

- The contribution of the second diagram is obtained exactly in the same way. The only difference with the first diagram is that the leg of electron 1 (resp. 2) in the final state is connected to the same vertex as the leg of electron 2 (resp. 1) in the initial state. It is easy to take this into account, obtaining

$$\frac{ie^2}{q'^2 + i\epsilon}[\bar{u}_{s_2'}(\vec{p}_2')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_1'}(\vec{p}_1')\gamma_\mu u_{s_2}(\vec{p}_2)]\,,$$

where now $q' = p_2' - p_1 = p_2 - p_1'$. Since the two diagrams differ by the exchange of an electron, which obeys Fermi-Dirac statistics, the two contributions must be *subtracted* from each other. More explicitly, one has that while for the first diagram one chooses the creation and annihilation operators at vertices 1 and 2 as follows,

$$\begin{array}{lcll} (\bar{\psi}\gamma^\mu\psi A_\mu)_1 & & (\bar{\psi}\gamma^\mu\psi A_\mu)_2 & \\ \hookrightarrow b(\vec{p}_1')^\dagger b(\vec{p}_1) & \times & \hookrightarrow b(\vec{p}_2')^\dagger b(\vec{p}_2) & = +b(\vec{p}_2')^\dagger b(\vec{p}_1')^\dagger b(\vec{p}_1)b(\vec{p}_2)\,, \end{array}$$

for the second diagram one has instead

$$\begin{array}{lcll} (\bar{\psi}\gamma^\mu\psi A_\mu)_1 & & (\bar{\psi}\gamma^\mu\psi A_\mu)_2 & \\ \hookrightarrow b(\vec{p}_2')^\dagger b(\vec{p}_1) & \times & \hookrightarrow b(\vec{p}_1')^\dagger b(\vec{p}_2) & = -b(\vec{p}_2')^\dagger b(\vec{p}_1')^\dagger b(\vec{p}_1)b(\vec{p}_2)\,. \end{array}$$

- The final result for M_{fi} to order $\mathcal{O}(e^2)$ is then

$$\begin{aligned} iM_{fi} = ie^2\Bigg\{ & \frac{1}{(p_1' - p_1)^2}[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_2'}(\vec{p}_2')\gamma_\mu u_{s_2}(\vec{p}_2)] \\ & - \frac{1}{(p_2' - p_1)^2}[\bar{u}_{s_2'}(\vec{p}_2')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_1'}(\vec{p}_1')\gamma_\mu u_{s_2}(\vec{p}_2)]\Bigg\}\,. \end{aligned} \tag{3.62}$$

One easily recognises that $(p_1' - p_1)^2 = t$ and $(p_2' - p_1)^2 = u$ are the usual Mandelstam variables.

Cross section for Møller scattering The scattering amplitude Eq. (3.62) is the starting point to obtain the electron-electron cross section to lowest order in the coupling e. This involves $|M_{fi}|^2$, and to obtain it notice that

$$[\bar{u}_{s'}(\vec{p}')\gamma^\mu u_s(\vec{p})]^* = u_s(\vec{p})^\dagger(\gamma^\mu)^\dagger(\gamma^0)^\dagger u_{s'}(\vec{p}') = \bar{u}_s(\vec{p})\gamma^0(\gamma^\mu)^\dagger(\gamma^0)^\dagger u_{s'}(\vec{p}')\,, \tag{3.63}$$

but since $\gamma^0\gamma^\mu\gamma^0 = (\gamma^\mu)^\dagger$ [check Eq. (3.14)], we have

$$[\bar{u}_{s'}(\vec{p}')\gamma^\mu u_s(\vec{p})]^* = \bar{u}_s(\vec{p})\gamma^\mu u_{s'}(\vec{p}')\,. \tag{3.64}$$

Using this result, we find for $|M_{fi}|^2$

$$\begin{aligned}|M_{fi}|^2 = e^4\Big\{&\frac{1}{t^2}[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_1}(\vec{p}_1)\gamma^\nu u_{s_1'}(\vec{p}_1')][\bar{u}_{s_2'}(\vec{p}_2')\gamma_\mu u_{s_2}(\vec{p}_2)][\bar{u}_{s_2}(\vec{p}_2)\gamma_\nu u_{s_2'}(\vec{p}_2')]\\ &+\frac{1}{u^2}[\bar{u}_{s_2'}(\vec{p}_2')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_2}(\vec{p}_2)\gamma^\nu u_{s_1'}(\vec{p}_1')][\bar{u}_{s_1'}(\vec{p}_1')\gamma_\mu u_{s_2}(\vec{p}_2)][\bar{u}_{s_1}(\vec{p}_1)\gamma_\nu u_{s_2'}(\vec{p}_2')]\\ &-\frac{2}{tu}\mathrm{Re}\,[\bar{u}_{s_2'}(\vec{p}_2')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_1}(\vec{p}_1)\gamma^\nu u_{s_1'}(\vec{p}_1')][\bar{u}_{s_1'}(\vec{p}_1')\gamma_\mu u_{s_2}(\vec{p}_2)][\bar{u}_{s_2}(\vec{p}_2)\gamma_\nu u_{s_2'}(\vec{p}_2')]\Big\}\,.\end{aligned} \tag{3.65}$$

This transition probability enters the cross section for a process in which the initial electron have a definite spin component, and the spin component of the final electrons is measured. It is more frequent in experiments to deal with initial beams of electrons whose spin is up or down with equal probability, and to not measure the spin of the final particles. The corresponding cross section involves then the average over $s_{1,2}$ and the same over $s'_{1,2}$. The relevant transition probability is then

$$\langle|M_{fi}|^2\rangle \equiv \frac{1}{2^2}\sum_{s_1,s_2,s_1',s_2'}|M_{fi}|^2\,. \tag{3.66}$$

To compute the sum in Eq. (3.66), we notice first that

$$[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_1}(\vec{p}_1)\gamma^\nu u_{s_1'}(\vec{p}_1')] = \mathrm{tr}\,\left(\gamma^\mu u_{s_1}(\vec{p}_1)\bar{u}_{s_1}(\vec{p}_1)\gamma^\nu u_{s_1'}(\vec{p}_1')\bar{u}_{s_1'}(\vec{p}_1')\right)\,, \tag{3.67}$$

and similarly for the other factors in the first two terms, and that

$$\begin{aligned}&[\bar{u}_{s_2'}(\vec{p}_2')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{u}_{s_1}(\vec{p}_1)\gamma^\nu u_{s_1'}(\vec{p}_1')][\bar{u}_{s_1'}(\vec{p}_1')\gamma_\mu u_{s_2}(\vec{p}_2)][\bar{u}_{s_2}(\vec{p}_2)\gamma_\nu u_{s_2'}(\vec{p}_2')]\\ &=\mathrm{tr}\,\left(\gamma^\mu u_{s_1}(\vec{p}_1)\bar{u}_{s_1}(\vec{p}_1)\gamma^\nu u_{s_1'}(\vec{p}_1')\bar{u}_{s_1'}(\vec{p}_1')\gamma_\mu u_{s_2}(\vec{p}_2)\bar{u}_{s_2}(\vec{p}_2)\gamma_\nu u_{s_2'}(\vec{p}_2')\bar{u}_{s_2'}(\vec{p}_2')\right)\,.\end{aligned} \tag{3.68}$$

We then use the identities Eq. (3.28) to write

$$\begin{aligned}\langle|M_{fi}|^2\rangle = \frac{e^4}{4}\Big\{&\frac{1}{t^2}\mathrm{tr}\,\left(\gamma^\mu(\not{p}_1+m)\gamma^\nu(\not{p}_1'+m)\right)\mathrm{tr}\,\left(\gamma_\mu(\not{p}_2+m)\gamma_\nu(\not{p}_2'+m)\right)\\ &+\frac{1}{u^2}\mathrm{tr}\,\left(\gamma^\mu(\not{p}_1+m)\gamma^\nu(\not{p}_2'+m)\right)\mathrm{tr}\,\left(\gamma_\mu(\not{p}_2+m)\gamma_\nu(\not{p}_1'+m)\right)\\ &-\frac{2}{tu}\mathrm{Re}\,\mathrm{tr}\,\left(\gamma^\mu(\not{p}_1+m)\gamma^\nu(\not{p}_1'+m)\gamma_\mu(\not{p}_2+m)\gamma_\nu(\not{p}_2'+m)\right)\Big\}\,.\end{aligned} \tag{3.69}$$

To compute the traces we need the following results, that follow from the basic relation Eq. (3.15):

$$\begin{aligned}
&\mathrm{tr}\,\gamma^{\mu_1}\dots\gamma^{\mu_{2n+1}} = 0\,,\\
&\mathrm{tr}\,\gamma^\mu\gamma^\nu = 4\eta^{\mu\nu}\,,\\
&\mathrm{tr}\,\gamma^\mu\gamma^\rho\gamma^\nu\gamma^\sigma = 4\left(\eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\mu\nu}\eta^{\rho\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho}\right)\,,\\
&\gamma^\mu \not{a}\gamma_\mu = -2\not{a}\,,\ \gamma^\mu\not{a}\not{b}\gamma_\mu = 4a\cdot b\,,\\
&\gamma^\mu\not{a}\not{b}\not{c}\gamma_\mu = -2\not{c}\not{b}\not{a}\,.
\end{aligned} \tag{3.70}$$

The first relation tells us that in the traces appearing in the first two terms, the only nonzero contributions come from terms where both momenta appear, or neither. For the first factor of the first term we find

$$\begin{aligned}
\mathrm{tr}\,\left(\gamma^\mu(\not{p}_1 + m)\gamma^\nu(\not{p}'_1 + m)\right) &= \mathrm{tr}\,\left(\gamma^\mu\not{p}_1\gamma^\nu\not{p}'_1\right) + m^2\mathrm{tr}\,\left(\gamma^\mu\gamma^\nu\right)\\
&= 4\left(p_1^\mu p_1'^\nu + p_1^\nu p_1'^\mu - \eta^{\mu\nu}\left(p_1\cdot p'_1 - m^2\right)\right)\,,
\end{aligned} \tag{3.71}$$

and similarly for the second factor. Contracting the two we find

$$\begin{aligned}
&\mathrm{tr}\,\left(\gamma^\mu(\not{p}_1 + m)\gamma^\nu(\not{p}'_1 + m)\right)\mathrm{tr}\,\left(\gamma_\mu(\not{p}_2 + m)\gamma_\nu(\not{p}'_2 + m)\right)\\
&= 16\left(p_1^\mu p_1'^\nu + p_1^\nu p_1'^\mu - \eta^{\mu\nu}\left(p_1\cdot p'_1 - m^2\right)\right)\left(p_{2\mu}p'_{2\nu} + p_{2\nu}p'_{2\mu} - \eta_{\mu\nu}\left(p_2\cdot p'_2 - m^2\right)\right)\\
&= 32\left(p_1\cdot p_2\, p'_1\cdot p'_2 + p_1\cdot p'_2\, p'_1\cdot p_2 + m^2(2m^2 - p_1\cdot p'_1 - p_2\cdot p'_2)\right)\\
&= 8\left((s-2m^2)^2 + (u-2m^2)^2 + 4m^2 t\right)\,,
\end{aligned} \tag{3.72}$$

where we used

$$\begin{aligned}
p_1\cdot p_2 &= \ \ \tfrac{1}{2}[(p_1+p_2)^2 - 2m^2] = \ \ \tfrac{1}{2}[s-2m^2] = p'_1\cdot p'_2\,,\\
p_1\cdot p'_1 &= -\tfrac{1}{2}[(p_1-p'_1)^2 - 2m^2] = -\tfrac{1}{2}[t-2m^2] = p_2\cdot p'_2\,,\\
p_1\cdot p'_2 &= -\tfrac{1}{2}[(p_1-p'_2)^2 - 2m^2] = -\tfrac{1}{2}[u-2m^2] = p_2\cdot p'_1\,.
\end{aligned} \tag{3.73}$$

The second term is found simply by exchanging p'_1 and p'_2, and so u and t. For the last term we find

$$\begin{aligned}
&\mathrm{tr}\,\left(\gamma^\mu(\not{p}_1 + m)\gamma^\nu(\not{p}'_1 + m)\gamma_\mu(\not{p}_2 + m)\gamma_\nu(\not{p}'_2 + m)\right)\\
&= \mathrm{tr}\,\left(\gamma^\mu\left(\not{p}_1\gamma^\nu\not{p}'_1 + m(\not{p}_1\gamma^\nu + \gamma^\nu\not{p}'_1) + m^2\gamma^\nu\right)\gamma_\mu(\not{p}_2 + m)\gamma_\nu(\not{p}'_2 + m)\right)\\
&= \mathrm{tr}\,\left(\left(-2\not{p}'_1\gamma^\nu\not{p}_1 + 4m(p_1+p'_1)^\nu - 2m^2\gamma^\nu\right)(\not{p}_2 + m)\gamma_\nu(\not{p}'_2 + m)\right)\\
&= -2\,\mathrm{tr}\,\left(\left(\not{p}'_1\gamma^\nu\not{p}_1 + m^2\gamma^\nu\right)\left(\not{p}_2\gamma_\nu\not{p}'_2 + m^2\gamma_\nu\right)\right) + 4m^2(p_1+p'_1)^\nu\,\mathrm{tr}\,\left(\not{p}_2\gamma_\nu + \gamma_\nu\not{p}'_2\right)\\
&= -8\left(4p_1\cdot p_2\, p_1\cdot p'_2 - 2m^2(p_2\cdot p'_2 + p_1\cdot p'_1) + 4m^4\right) + 16m^2(p_1+p'_1)\cdot(p_2+p'_2)\\
&= -32\left(p_1\cdot p_2\, p_1\cdot p'_2 - \tfrac{1}{2}m^2(p_1\cdot p_2 + p_1\cdot p'_2 + p'_1\cdot p_2 + p'_1\cdot p'_2 + p_2\cdot p'_2 + p_1\cdot p'_1) + m^4\right)\\
&= -8\left((s-2m^2)^2 + 2m^2(t+u-s)\right)\,.
\end{aligned} \tag{3.74}$$

Plugging everything in Eq. (3.69) we find

$$\begin{aligned}
\langle|M_{fi}|^2\rangle = 2e^4\Bigg\{&\frac{1}{t^2}\left((s-2m^2)^2 + (u-2m^2)^2 + 4m^2 t\right)\\
&+ \frac{1}{u^2}\left((s-2m^2)^2 + (t-2m^2)^2 + 4m^2 u\right)\\
&+ \frac{2}{tu}\left((s-2m^2)^2 + 2m^2(t+u-s)\right)\Bigg\}\,.
\end{aligned} \tag{3.75}$$

To obtain the cross section we need the formula

$$d\sigma = \frac{\langle|M_{fi}|^2\rangle}{4\sqrt{(p_1\cdot p_2)^2-m^4}}d\Phi^{(2)} = \frac{\langle|M_{fi}|^2\rangle}{2\sqrt{s(s-4m^2)}}\frac{d\Omega_{\rm CM}}{32\pi^2}\sqrt{1-\frac{4m^2}{s}} = \frac{\langle|M_{fi}|^2\rangle}{64\pi^2 s}d\Omega_{\rm CM}\,. \tag{3.76}$$

We find

$$\begin{aligned}\frac{d\sigma}{d\Omega_{\rm CM}} = \frac{e^4}{32\pi^2 s}\Bigg\{&\frac{1}{t^2}\left((s-2m^2)^2+(u-2m^2)^2+4m^2t\right)\\ &+\frac{1}{u^2}\left((s-2m^2)^2+(t-2m^2)^2+4m^2u\right)\\ &+\frac{2}{tu}\left((s-2m^2)^2+2m^2(t+u-s)\right)\Bigg\}.\end{aligned} \tag{3.77}$$

This expression simplifies considerably in the ultrarelativistic limit where momenta are much larger than masses of the particles. In this limit we can neglect higher orders in m, and get

$$\begin{aligned}\left.\frac{d\sigma}{d\Omega_{\rm CM}}\right|_{\rm ultrarel} &= \frac{e^4}{32\pi^2 s}\left\{\frac{s^2+u^2}{t^2}+\frac{s^2+t^2}{u^2}+\frac{2s^2}{tu}\right\}\\ &= \frac{e^4}{32\pi^2 s}\left\{\frac{(s^2+u^2)u^2+(s^2+t^2)t^2+2s^2tu}{t^2u^2}\right\}\\ &= \frac{e^4}{32\pi^2 s}\left\{\frac{s^2(u+t)^2+t^4+u^4}{t^2u^2}\right\} = \frac{e^4}{32\pi^2 s}\frac{s^4+t^4+u^4}{t^2u^2}\,,\end{aligned} \tag{3.78}$$

since $s+t+u=4m^2\simeq 0$. In the centre of mass

$$t = -\frac{s}{2}(1-\cos\theta)\,,\qquad u = -\frac{s}{2}(1+\cos\theta)\,, \tag{3.79}$$

and so

$$\begin{aligned}\left.\frac{d\sigma}{d\Omega_{\rm CM}}\right|_{\rm ultrarel} &= \frac{\alpha^2}{2s}\frac{16+(1-\cos\theta)^4+(1-\cos\theta)^4}{(1-\cos^2\theta)^2} = \frac{\alpha^2}{2s}\frac{18+12\cos^2\theta+2\cos^4\theta}{\sin^4\theta^2}\\ &= \frac{\alpha^2}{s}\frac{(3+\cos^2\theta)^2}{\sin^4\theta^2}\,,\end{aligned} \tag{3.80}$$

where $\alpha = e^2/(4\pi)$ is the fine structure constant.

Bhabha scattering We then proceed to discuss electron-positron scattering (*Bhabha scattering*), see Fig. 14. The left panel of Fig. 14 corresponds to electron-positron interaction mediated by the exchange of a photon. This diagram is identical to the left diagram in Fig. 13 after reflecting the rightmost fermion line upside down. This is the graphical device employed to represent an antifermion. The right panel instead corresponds to a different way of interacting, with the electron and positron annihilating into a photon, which subsequently undergoes pair production. Let us evaluate these diagrams.

- The first diagram is evaluated exactly as above. The leftmost fermion line corresponding to the electron reads as before,

$$\bar{u}_{s_1'}(\vec{p}_1')(-ie)\gamma^\mu u_{s_1}(\vec{p}_1)\,.$$

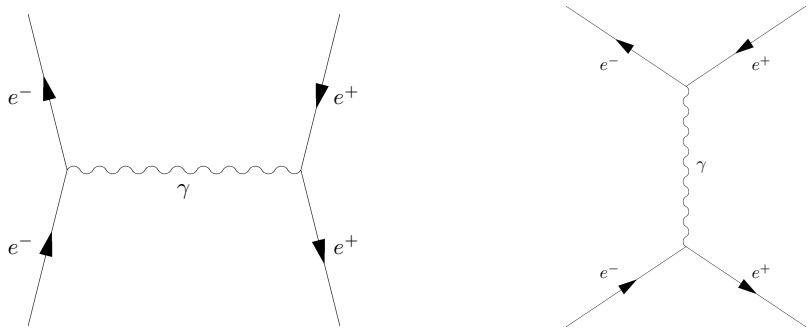

Figure 14: Bhabha scattering.

- For the rightmost line, following our prescription of starting from the end of a fermion line requires that we begin with the wave function of the positron in the initial state. This is

$$\bar{v}_{s_2}(\vec{p}_2)\,.$$

  Following the line backwards we encounter the second interaction vertex, and so we include a factor

$$-ie\gamma^\nu\,.$$

  We finally meet the external leg of the positron in the final state, and so include the corresponding wave function,

$$v_{s_2'}(\vec{p}_2')\,.$$

- The internal photon lines contributes again a factor

$$D_{\mu\nu}(q) = -i\frac{\eta_{\mu\nu}}{q^2+i\epsilon}\,,$$

  with $q = p_1' - p_1 = p_2 - p_2'$. Putting everything together we find

$$\frac{ie^2}{q^2+i\epsilon}[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{v}_{s_2}(\vec{p}_2)\gamma_\mu v_{s_2'}(\vec{p}_2')]\,.$$

- Since an antifermion line crosses the diagram all the way from the final to the initial state, an extra minus sign is required. This is understood again in terms of pairing vertices with incoming and outgoing particles: since $\bar{\psi}$ (resp. ψ) annihilates (resp. creates) an antifermion, we have

$$\begin{array}{lcll} (\bar{\psi}\gamma^\mu\psi A_\mu)_1 & & (\bar{\psi}\gamma^\mu\psi A_\mu)_2 & \\ \hookrightarrow b(\vec{p}_1')^\dagger b(\vec{p}_1) & \times & \hookrightarrow d(\vec{p}_2)d(\vec{p}_2')^\dagger & = -d(\vec{p}_2')^\dagger b(\vec{p}_1')^\dagger b(\vec{p}_1)d(\vec{p}_2)\,, \end{array}$$

- The second diagram is evaluated again according to the same prescriptions, only this time the fermion lines both begin and end in the initial or final state. No extra sign is required since

$$\begin{array}{lcll} (\bar{\psi}\gamma^\mu\psi A_\mu)_1 & & (\bar{\psi}\gamma^\mu\psi A_\mu)_2 & \\ \hookrightarrow d(\vec{p}_2)b(\vec{p}_1) & \times & \hookrightarrow b(\vec{p}_1')^\dagger d(\vec{p}_2')^\dagger & = +d(\vec{p}_2')^\dagger b(\vec{p}_1')^\dagger b(\vec{p}_1)d(\vec{p}_2)\,. \end{array}$$

- Start from the line in the final state, following it backwards. According to the rules that we have already formulated, one requires to include a final-state electron wave function,

$$\bar{u}_{s_1'}(\vec{p}_1')\,,$$

a vertex factor,

$$-ie\gamma^\mu\,,$$

and a final-state positron wave function,

$$v_{s_2'}(\vec{p}_2')\,,$$

resulting into

$$\bar{u}_{s_1'}(\vec{p}_1')(-ie)\gamma^\mu v_{s_2'}(\vec{p}_2')\,.$$

- For the line in the initial state, we write down the factors as follows: an initial-state positron wave function,

$$\bar{v}_{s_2}(\vec{p}_2)\,,$$

a vertex factor,

$$-ie\gamma^\nu\,,$$

and an initial-state electron wave function,

$$u_{s_1}(\vec{p}_1)\,,$$

resulting into

$$\bar{v}_{s_2}(\vec{p}_2)(-ie)\gamma^\nu u_{s_1}(\vec{p}_1)\,.$$

- We include now the familiar factor for an internal photon line,

$$D_{\mu\nu}(q') = -i\frac{\eta_{\mu\nu}}{q'^2 + i\epsilon}\,,$$

where now $q' = p_1 + p_2 = p_1' + p_2'$ is the total four-momentum of the system. The final expression is

$$\frac{ie^2}{q'^2 + i\epsilon}[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu v_{s_2'}(\vec{p}_2')][\bar{v}_{s_2}(\vec{p}_2)\gamma_\mu u_{s_1}(\vec{p}_1)]\,.$$

- The full result for $\mathcal{O}(e^2)$ Bhabha scattering is

$$iM_{fi} = -ie^2\Bigg\{\frac{1}{(p_1' - p_1)^2}[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu u_{s_1}(\vec{p}_1)][\bar{v}_{s_2}(\vec{p}_2)\gamma_\mu v_{s_2'}(\vec{p}_2')] - \frac{1}{(p_1 + p_2)^2}[\bar{u}_{s_1'}(\vec{p}_1')\gamma^\mu v_{s_2'}(\vec{p}_2')\bar{v}_{s_2}(\vec{p}_2)\gamma_\mu u_{s_1}(\vec{p}_1)]\Bigg\}\,.$$

This can now be used to obtain the scattering cross section exactly as we did above for Møller scattering. We report here only the final result for the differential criss section summed over final spins and averged over initial spins in the ultrarelativistic limit,

$$\left.\frac{d\sigma}{d\Omega_{\rm CM}}\right|_{\rm ultrarel} = \frac{\alpha^2}{s}\frac{(3 + \cos^2\theta)^2}{(1 - \cos\theta)^2} = \frac{\alpha^2}{s}\frac{(3 + \cos^2\theta)^2}{4\sin^4\frac{\theta}{2}}\,. \tag{3.81}$$

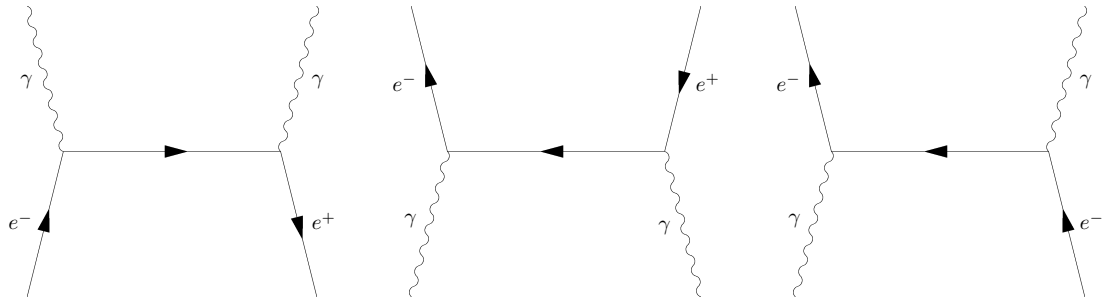

Figure 15: Left: e^-- e^+ annihilation. Centre: pair production. Right: Compton scattering.

There are five more processes described by combining two electromagnetic vertices. Diagrams for positron-positron scattering are obtained from Fig. 13 by flipping both fermion lines, and the corresponding scattering amplitude can be obtained using only the rules already discussed above. In practice, the wave functions $u_s(\vec{p})$ and $\bar{u}_{s'}(\vec{p}')$ for initial and final fermions are replaced by $v_{s'}(\vec{p}')$ and $\bar{v}_s(\vec{p})$ for final and initial antifermions. Diagrams for other three processes are shown in Fig. 15, and correspond (from left to right) to *electron-positron annihilation*, *electron-positron pair creation*, and *Compton scattering* of photons on electrons. Notice that the internal line corresponds now to a virtual fermion, and is oriented consistently with the external lines. (For each of these processes there is a second diagram: can you draw it?) Notice also that *two* photons appear in the physical annihilation and pair-creation processes, as necessary for energy-momentum conservation. Finally, Compton scattering of photons on positrons can also take place, and the relevant diagrams are obtained straighforwardly from those for photon-electron scattering. (Can you draw them?)

These diagrams require two further Feynman rules, to deal with external photon lines and internal fermion lines. Let us discuss the rightmost diagram in Fig. 15, corresponding to Compton scattering.

- Start again from the end of the fermion line and proceed backwards. Using the known rules, we write a final-state electron wave function and a vertex factor, getting

$$\bar{u}_{s'_1}(\vec{p}'_1)(-ie)\gamma^\mu \,.$$

- We now need the appropriate factor for an internal fermion line. This is

$$S(q) = i\frac{\not{q} + m}{q^2 - m^2 + i\epsilon}\,,$$

  where $\not{q} = q_\mu\gamma^\mu$ with q the momentum carried by the virtual internal fermion, flowing in the same direction as the arrow. By energy-momentum conservation, $q = \vec{p}'_1 - k_1$. This should be appended to the expression above, followed by the other vertex factor and by initial-state electron wave function. This results in

$$\bar{u}_{s'_1}(\vec{p}'_1)(-ie)\gamma^\mu i\frac{\not{q} + m}{q^2 - m^2 + i\epsilon}(-ie)\gamma^\nu u_{s_1}(\vec{p}_1)\,.$$

- The external photon legs require only the inclusion of suitable photon wave functions. For circularly polarised photons, these are

$$\epsilon^\lambda_\mu(k_1)\,, \qquad \epsilon^{\lambda'}_\nu(k'_1)^*$$

for the initial and final leg, respectively. Here $\lambda = \pm$ is the helicity of the photon, $k = (|\vec{k}|, \vec{k})$, and setting $\vec{k} = |\vec{k}|\hat{n}$ for the photon momentum, we have

$$\epsilon_\mu^\lambda(k) = \frac{1}{\sqrt{2}} R(\hat{n})(0, 1, i\lambda, 0)\,,$$

where $R(\hat{n})$ is the rotation that bring the third axis to the direction $\hat{n}$.

- The final result is

$$-ie^2 \bar{u}_{s_1'}(\vec{p}_1') \not{\epsilon}^\lambda(k_1) \frac{\not{q} + m}{q^2 - m^2 + i\epsilon} \not{\epsilon}^{\lambda'}(k_1')^* u_{s_1}(\vec{p}_1)\,.$$

A few more comments.

- The only relevant aspect distinguishing two diagrams contributing to the same process is their topology. Consider then the diagram of Fig. 16: does it represent a different microscopic process for elastic electron-positron scattering than the one in the right panel of Fig. 14, and so should one include also its contribution in the scattering amplitude? The answer is no: all that matters is the *topology* of the diagram, and of course the content of the initial and final states as encoded in the external lines, i.e., the "before" and "after" of the process, which determine what kind of reaction we are considering. When exactly the annihilation of the first pair happens, and when the creation of the second pair happens, are actually meaningless questions, and should not (and do not) matter when listing all the different ways in which a given process is realised at the microscopic level.

- Certain diagrams require the inclusion of suitable symmetry factors, whise origin can be understood as follows. The diagrams depicted above originate from the contraction of the lines stemming out of the various vertices, among themselves and with the external lines corresponding to the incoming and outgoing particles. One should count in how many ways these contractions can be made for a given diagram topology, and include the factor $1/n!$ from the expansion of Dyson's formula, to obtain the correct factor.

- In higher perturbative orders, internal lines may form loops, see Fig. 17. In this case the four-momentum of the internal lines is not entirely fixed by conservation at each vertex. One deal with this by assigning a momentum to each internal line and including a momentum-conserving delta function for each vertex; after using one of them to express overall conservation of energy and momentum, and integrating over the momenta of the internal lines, some internal momenta will be fixed by the delta functions, while for the others one needs to perform the remaining integrations.[43] In addition, each fermion loop requires the inclusion of an extra minus sign.

- Diagrams differing by the exchange of external boson lines should be added to each other; those differing by the exchange of external fermion lines, as already pointed out above, should be subtracted from each other.

The left panel in Fig. 17 shows the simplest contribution to *light-by-light (Delbrück) scattering*, while the right panel corresponds to a more complicated way in which electron-electron

[43] This is much less straightforward than it seems, leads to one of the most disturbing aspects of QFT - that of ultraviolet divergences. We will not get any closer to this problem than this brief remark.

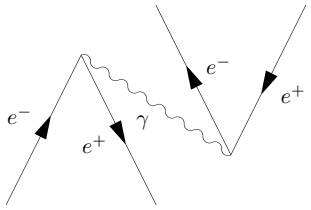

Figure 16: The same diagram as in the right panel of Fig. 14, only drawn in a different way.

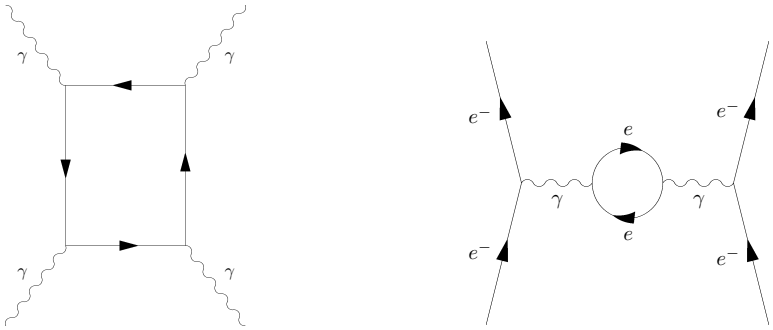

Figure 17: More complicated diagrams.

scattering can take place, with the exchanged photon turning temporarily into an electron-positron pair. As already mentioned above, the contribution of a Feynman diagram to the transition amplitude contains a coupling constant factor for each vertex, so a factor of e for each electromagnetic vertex. The diagrams of Figs. 13, 14, and 15 are all proportional to $e^2 = 4\pi\alpha$ (in natural units), with $\alpha \simeq 1/137$ the *fine structure constant*. The diagrams of Fig. 17 are proportional instead to α^2, and so the contribution to elastic electron-electron scattering of the diagram in the right panel is relatively suppressed with respect to those in Fig. 13. In general, to describe a given process to a given precision we will then need only a limited number of diagrams, as the introduction of more interaction vertices will further suppress the contribution of the diagram. Moreover, a process like light-by-light scattering has overall a smaller transition amplitude than any of the processes depicted in Figs. 13, 14, and 15. This is consistent with the fact that this kind of process is not expected at the level of classical electrodynamics.

Strong interactions Feynman rules for processes involving quarks and gluons differ from those found for electromagnetic interactions for the appearance of colour, and for the presence of further vertices coupling the gluons to themselves. For instance, the quark-quark-gluon vertex reads

$$-ig\gamma^\mu t^a_{ij}\,,$$

where t^a, $a = 1, \ldots, 8$, is one of the eight linearly independent 3×3 complex Hermitian traceless matrices, with indices i, j corresponding to the colour states of the quarks, and the index a corresponding to the colour states of the gluon. The matrices t^a are usually chosen with normalisation $\mathrm{tr}\, t^a t^b = \frac{1}{2}\delta^{ab}$. Similarly, the gluon propagator differs from the photon propagator for the presence of an extra factor δ^{ab}, basically carrying colour from one vertex to the other,

$$D^{ab}_{g\,\mu\nu}(q) = -i\frac{\eta_{\mu\nu}\delta^{ab}}{q^2 + i\epsilon}\,.$$

The lowest-order amplitude for quark-quark scattering is obtained by the analogues of the diagrams in Fig. 13 obtained replacing the electron lines by quark lines, and the photon lines by gluon lines. The result is

$$iM_{fi} = ig^2 \Big\{ \frac{1}{(p_1' - p_1)^2} [\bar{u}^{c_1'}_{s_1'}(\vec{p}_1')\gamma^\mu t^a u^{c_1}_{s_1}(\vec{p}_1)][\bar{u}^{c_2'}_{s_2'}(\vec{p}_2')\gamma_\mu t^a u^{c_2'}_{s_2}(\vec{p}_2)] \\ - \frac{1}{(p_2' - p_1)^2} [\bar{u}^{c_2'}_{s_2'}(\vec{p}_2')\gamma^\mu t^a u^{c_1}_{s_1}(\vec{p}_1)][\bar{u}^{c_1'}_{s_1'}(\vec{p}_1')\gamma_\mu t^a u^{c_2}_{s_2}(\vec{p}_2)] \Big\} , \tag{3.82}$$

where $c_{1,2}$ and $c'_{1,2}$ are the colours of the initial and final quarks, and summation over the repeated index a is understood. The calculation of cross section summed over final colours and averaged over initial colours is simplified by the following identities,

$$\sum_{s,c} [u^c_s(\vec{p})]_i [\bar{u}^c_s(\vec{p})]_j = (\not{p} + m)\delta_{ij} , \qquad \sum_a t^a_{ij} t^a_{kl} = \frac{1}{2}\left(\delta_{il}\delta_{jk} - \frac{1}{3}\delta_{ij}\delta_{kl}\right) , \tag{3.83}$$

where i, j are the colour indices of the quark wave functions. While this looks quite similar to the electromagnetic case (up to the extra colour degree of freedom), the gluon self-interaction vertices lead to a completely different physics, with quarks and gluons not even appearing as physical states in scattering experiments. Nonetheless, the use of perturbation theory and of the particle-exchange description can be used reliably at high energies.

Weak interactions Let us consider now processes mediated by the weak interactions. The simplest such process is the main decay mode of a muon (Fig. 18, left), $\mu^- \to e^- \nu_\mu \bar{\nu}_e$ (notice the separate conservation of the electronic and the muonic lepton numbers). Except for the presence of spectator quarks that do not take part in the decay process, the same type of diagram describes the beta decay of the neutron, $n \to p\, e^- \bar{\nu}_e$ (Fig. 18, centre). The diagram for the process of antineutrino capture, $\bar{\nu}_e\, p \to e^+ n$, is obtained from this one by switching the interacting u and d quarks, replacing the electron with a positron, and moving the antineutrino to the initial state. Finally, again essentially the same diagram, except with electron and electron antineutrino replaced by antimuon and muonic neutrino, describes also the main decay mode of the positively charged pion (Fig. 18, right).

The new rules required here are those for the interaction vertex and for the propagator of the W boson. These read respectively

$$-i\frac{g_W}{\sqrt{2}}\gamma^\mu \frac{1-\gamma^5}{2} ,$$

times the appropriate CKM matrix element if quarks are involved and

$$D_{W\,\mu\nu}(q) = -i\frac{\eta_{\mu\nu} - \frac{q_\mu q_\nu}{m_W^2}}{q^2 - m_W^2 + i\epsilon} .$$

Rules for the other interaction vertices will not be discussed. The propagator of the Z boson is identical to D_W except for replacing m_W with m_Z.

The interaction vertex is similar to the one found in the electromagnetic case, with two important differences:

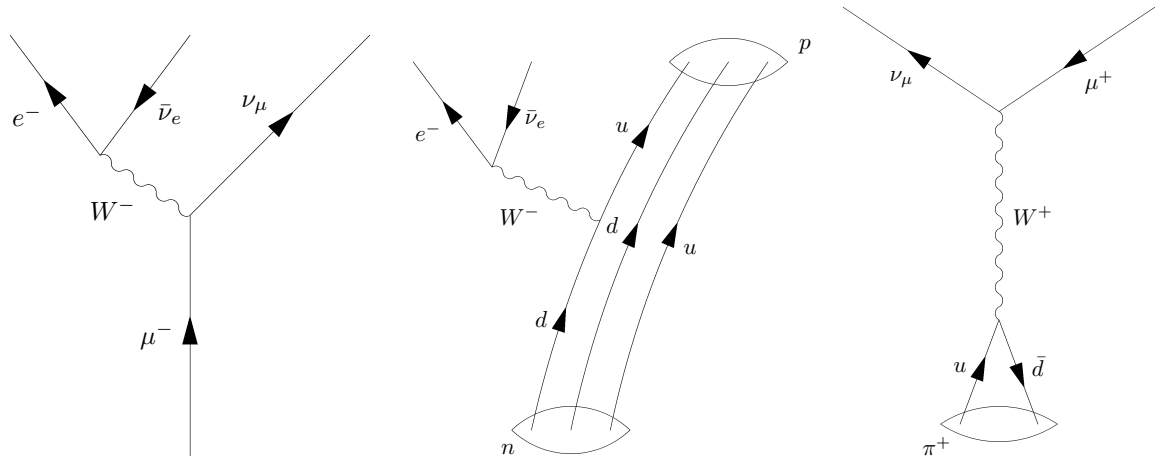

Figure 18: Left: muon decay. Centre: Beta decay of the neutron. Right: Charged pion decay.

- it changes the type of fermion (lepton or quark);
- it includes a projector $\frac{1-\gamma^5}{2}$ that selects only one of the chiralities of the fermions.

The second point is the reason behind parity violations in weak interactions (more below). While the individual leptonic numbers are not conserved anymore, it is clear that lepton family number still are. On the other hand, since quark flavours are mixed by the CKM matrix in weak processes, the analogous "quark family" number is not conserved. Still, the identification of quark families makes sense since the diagonal elements of the CKM matrix are the largest ones, meaning that u, c, t are more strongly coupled respectively to d, s, b by weak interactions.

As we have already discussed, for processes like those in Fig. 18, where the momentum flowing in the internal line is much smaller than the W mass, one can approximate

$$D_{W\,\mu\nu}(q) \underset{q^2 \ll m_W^2}{\to} i\frac{\eta_{\mu\nu}}{m_W^2}\,.$$

The practical effect of this is that the two vertices connected by the propagator join together, leading to a point-like four-fermion interaction vertex, with effective coupling given by the Fermi constant $G_F = \frac{1}{\sqrt{2}}\left(\frac{g_w}{2M_W}\right)^2 \simeq 1.1 \cdot 10^{-5}$ GeV^{-2}. In this approximation, evaluating the relevant Feynman diagrams requires only to pair properly the wave functions of the incoming and outgoing fermions (including the gamma matrices attached to the vertices).

The description of weak interactions in terms of a four-fermion vertex appeared in the first theory of β decays proposed by Fermi in 1933. In this theory proton, neutron, electron and neutrino interacted directly via a four-fermion vertex, with coupling constant G_F. The fact that G_F has dimensions of some inverse power of mass indicated (for reasons that we will not discuss here) that the four-fermion interaction could not be the final word, and that a "more fundamental" theory of weak interactions had to exist (although it took forty years to find it). Notice that combining the measurement of G_F and assuming that the weak coupling is comparable to the electromagnetic one (as one would expect assuming unification) one finds $m_W \sim \sqrt{4\pi\alpha/G_F} \sim 90\,\text{GeV}$, which is correct to within 10%.

4 Symmetries

In this section we discuss in some detail the concept of symmetry in a physical theory and its consequences. Quoting S. Weinberg almost verbatim [7], a *symmetry* is a change in the experimenter's point of view that does not change the results of possible experiments. This statement seems quite cryptic at a first reading, so let us explain it in more detail. Consider two experimenters $\mathcal{O}$ and $\mathcal{O}'$ making measurements on the same physical system. Although they subscribe to the same operative rules concerning the measurement of the various observables, they use in general different reference objects to perform these measurements. For example, to measure the Cartesian coordinates of an object one needs a reference point (i.e., the origin of the coordinate system) and three orthogonal reference directions (i.e., the axes). The set of reference objects constitutes the *reference frame* of an observer. The use of different reference frames by $\mathcal{O}$ and $\mathcal{O}'$ generally leads to obtaining different values for the various physical quantities that they measure, and so to two different descriptions of the same system. However, if these descriptions lead to the same physical picture of the world, then the distinction is unimportant and the two descriptions are equivalent. In order for this to happen, the full set of outcomes that $\mathcal{O}$ and $\mathcal{O}'$ can obtain by carrying out any conceivable experiments must be the same: this is a first level of reading of Weinberg's definition of symmetry, taking "the results of possible experiments" to mean the full set of possible experimental results obtained by an experimenter. Also, a somewhat technical, "zeroth-level" reading is that if $\mathcal{O}$ sees two experimental outcomes as different, then so must do $\mathcal{O}'$, for otherwise a change in "point of view" would radically change the conclusions of the two experimenters. A second level of reading is that if two experimenters using different reference frames (and so a different "point of view") carry out the same experiment, with the same initial conditions, then they must obtain the same results. This requires that the dynamical laws governing the evolution of physical systems (that the experiments aim to unveil) must be the same in the two reference frames.

Summarising, the descriptions provided by $\mathcal{O}$ and $\mathcal{O}'$ using their respective reference frames are *equivalent* if (1) the set of all possible experimental outcomes, and (2) the dynamical laws governing the temporal evolution of physical systems are the same for the two experimenters. We add to this list the somewhat implicit request that (0) if $\mathcal{O}$ sees two states of the system as different, then so must do $\mathcal{O}'$. Identifying the observers $\mathcal{O}$ to $\mathcal{O}'$ with their reference frames, one then says that $\mathcal{O}$ and $\mathcal{O}'$ are equivalent, and that the change from $\mathcal{O}$ to $\mathcal{O}'$ is a symmetry, that leaves the physics unchanged. Another way to state the equivalence of two reference frames is two take one experimenter, place them in either of the reference frames $\mathcal{O}$ or $\mathcal{O}'$, and let them carry out whatever experiment they want. If the two frames are equivalent, then the experimenter will not be able to tell, neither by the mere results of the experiments nor by the physical laws that they can infer, in which of the two frames they have been placed.[44]

Let us now formalise the discussion above in mathematical terms. According to quantum mechanics, an experimenter can fully characterise the state of a physical system by carrying out measurements of a complete set of compatible observables. For example, in the case of a set of free particles these are the three components $\vec{p}$ of their momenta and one component of their spin, usually taken to be s_z. With the information collected in the experiments, an experimenter then assigns a suitable representative vector in the Hilbert space of the system, in correspondence with the values obtained for the various observables. Technically, this means that the expectation values of the operators corresponding to the various observables, computed on this vector, are

[44] Sometimes they can tell by the quality of the coffee.

equal to the values obtained experimentally. Now, two experimenters in reference frames $\mathcal{O}$ and $\mathcal{O}'$ generally find different experimental values; since the operators corresponding to the physical observables are the *same* for both observers (reflecting the fact that they operatively define an observable in the same way), the vectors that they assign to the same system are different. More precisely, since the overall phase and the norm of the representative vectors are unobservable, $\mathcal{O}$ and $\mathcal{O}'$ assign respectively the *rays*[45] $\mathcal{R}$ and $\mathcal{R}'$ to the state of the system, and these are generally different, and possibly belonging to different spaces of rays. However, if $\mathcal{O}$ and $\mathcal{O}'$ are equivalent, then the set of possible physical states that they can observe must be the same. Mathematically, this means that the Hilbert space $\mathcal{H}$ and the associated space of rays $\underline{\mathcal{H}}$, corresponding to the possible states that they can assign to the system, must be the same for $\mathcal{O}$ and $\mathcal{O}'$.

Establishing a relation between the two descriptions corresponds to defining a mapping $\mathcal{M}$ from the space of rays $\underline{\mathcal{H}} = \{\mathcal{R}\}$ to itself. To each ray observed by $\mathcal{O}$ there corresponds one and only one ray observed by $\mathcal{O}'$. Moreover, every ray corresponds to a possible observation of $\mathcal{O}'$, so the mapping must be surjective (onto). Finally, diffent rays must be mapped into different rays, since observing that two states of the system are different does not depend on the observer. These mathematical properties follow from the physical properties (0) and (1). We then have a mapping $\mathcal{M}$,

$$\begin{aligned} \mathcal{M} : \underline{\mathcal{H}} &\to \underline{\mathcal{H}} \\ \mathcal{R} &\mapsto \mathcal{R}' = \mathcal{M}\mathcal{R}\,, \end{aligned} \tag{4.1}$$

which is injective (one-to-one) and surjective, and therefore an invertible mapping. In a more direct way: if the two observers are equivalent, and there is a map from $\mathcal{O}$ to $\mathcal{O}'$, then there must also be an inverse mapping from $\mathcal{O}'$ to $\mathcal{O}$, for otherwise the two frames would be distinguishable and therefore not equivalent.

Representation of symmetries and Wigner's theorem Suppose that we now perform experiments on the system, and we see it transition from a state to another. Collecting the results of their experiments, the two observers will see that the transition between a given initial state, corresponding respectively to rays $\mathcal{R}_i$ and $\mathcal{R}'_i$, to a prescribed final state, corresponding respectively to rays $\mathcal{R}_f$ and $\mathcal{R}'_f$,

$$\mathcal{O} : \mathcal{R}_i \longrightarrow \mathcal{R}_f\,, \qquad \mathcal{O}' : \mathcal{R}'_i \longrightarrow \mathcal{R}'_f\,, \tag{4.2}$$

occurs with probabilities P and P',

$$P = (\mathcal{R}_i \cdot \mathcal{R}_f)^2\,, \qquad P' = (\mathcal{R}'_i \cdot \mathcal{R}'_f)^2\,, \tag{4.3}$$

where[46]

$$\mathcal{R}_1 \cdot \mathcal{R}_2 = |(\psi_1, \psi_2)|\,, \tag{4.4}$$

[45] A ray is an equivalence class of vectors in a Hilbert space $\mathcal{H}$ with respect to the equivalence relation "differing by a complex factor". Formally, given a vector $\psi \in \mathcal{H}$, the corresponding ray $\mathcal{R}_\psi \equiv [\psi]_\sim \in \mathcal{H}_{/\sim} \equiv \underline{\mathcal{H}}$ is the equivalence class of ψ with respect to the equivalence relation: $\psi \sim \phi$ if $\psi = c\phi$, $c \in \mathbb{C}$. Since transition probability between states represented by $\psi_{1,2}$ is $P_{12} = |(\psi_2, \psi_1)|^2[(\psi_2, \psi_2)(\psi_1, \psi_1)]^{-1}$, any choice of vectors in $[\psi_{1,2}]_\sim$ gives the same result.

[46] We use the convention that the scalar product (ψ, ϕ) is linear in ϕ and antilinear in ψ, i.e., $(\psi, \alpha\phi + \beta\phi') = \alpha(\psi, \phi) + \beta(\psi, \phi')$ and $(\alpha\psi + \beta\psi', \phi) = \alpha^*(\psi, \phi) + \beta^*(\psi', \phi)$.

with $\psi_{1,2}$ any normalised vector belonging to $\mathcal{R}_{1,2}$. Since $\mathcal{O}$ and $\mathcal{O}'$ are looking at the same physical processes, the transition probabilities they observe must be the same, $P = P'$, and so

$$\mathcal{R}_i \cdot \mathcal{R}_f = \mathcal{R}'_i \cdot \mathcal{R}'_f \,. \tag{4.5}$$

Since $\mathcal{R}'_{i,f} = \mathcal{M}\mathcal{R}_{i,f}$, we have that

$$\mathcal{R}_i \cdot \mathcal{R}_f = (\mathcal{M}\mathcal{R}_i) \cdot (\mathcal{M}\mathcal{R}_f) \,. \tag{4.6}$$

A theorem due to Wigner guarantees that any invertible transformation $\mathcal{M}$ on the space of rays $\underline{\mathcal{H}}$, $\mathcal{M} : \underline{\mathcal{H}} \to \underline{\mathcal{H}}$ that conserves probabilities (in the sense of Eq. (4.6)) can be implemented as a transformation M on the space of vectors $\mathcal{H}$, $M : \mathcal{H} \to \mathcal{H}$, that is either *linear and unitary* or *antilinear and antiunitary*. In other words, any such $\mathcal{M}$ can be obtained as the natural mapping between rays following from a unitary or antiunitary mapping M between vectors. Wigner's theorem implies that, without loss of generality, we can restrict our search for symmetry transformations, relating two equivalent observers, looking only at the set of unitary and antiunitary mappings of the Hilbert space of the system onto itself. More generally, Wigner's theorem applies when the two observers see the same space of possible states for the system under consideration, even if the dynamical laws they use to predict its temporal evolution are different. For a proof of the theorem, see Ref. [7].

We have already discussed linear unitary operators above (see p. 26). Antilinear antiunitary operators T have analogous properties: antilinearity means that for arbitrary $\psi, \phi \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{C}$,

$$T(\alpha\psi + \beta\phi) = \alpha^* T\psi + \beta^* T\phi \,. \tag{4.7}$$

An antiunitary operator is an antilinear operator that is norm-preserving, $\|T\psi\| = \|\psi\|$, and onto, i.e., having the whole of $\mathcal{H}$ as its image. Introducing the adjoint of an antilinear operator thourgh the definition

$$(\psi, T\phi) = (T^\dagger\psi, \phi)^* \,, \tag{4.8}$$

the two properties above are expressed respectively as $T^\dagger T = \mathbf{1}$ and $TT^\dagger = \mathbf{1}$. An antiunitary operator is then equivalently characterised as an antilinear operator satisfying $T^\dagger T = TT^\dagger = \mathbf{1}$.

Symmetries and temporal evolution What we discussed so far were the kinematical aspects of a symmetry, but we demand more from it: we want the physics to be the same for both observers. We then need that the dynamical evolution of a physical system be governed by the same laws for both observers, or stated differently we want that the equations of motion have the same form for both observers, i.e., they are invariant in form. If the dynamical evolution of the system is the same for both observers, it follows that the transformed of the evolved is equal to the evolved of the transformed: this entails that the Hamiltonian of the system is the same for both observers, and so is the physics.

In general, once we have established that the descriptions of the same physical system given by two observers are connected by a unitary or antiunitary transformation M, then we can relate also the dynamical laws that they use to predict the temporal evolution of the system. Let us assume for simplicity that the transformation does not involve time in any way, being time-independent and leaving untouched the time coordinate, and moreover let H and H' be the Hamiltonians used by observer $\mathcal{O}$ and $\mathcal{O}'$, respectively, which we again assume to be time-independent. If $\psi(0)$ is the state vector at $t = 0$ for the observer $\mathcal{O}$, at time t they will find

$$\psi(t) = U(t)\psi(0) = e^{-iHt}\psi(0) \,, \tag{4.9}$$

where $U(t) = e^{-iHt}$ is the (unitary) time-evolution operator. Similarly, for the observer $\mathcal{O}'$ at time $t = 0$ the state vector of the system is $\psi'(0)$, related to $\psi(0)$ by $\psi'(0) = M\psi(0)$, while at time t they see

$$\psi'(t) = U'(t)\psi'(0) = e^{-iH't}\psi'(0) = e^{-iH't}M\psi(0)\,. \tag{4.10}$$

On the other hand, connecting the descriptions of the two observers directly at time t one finds

$$\psi'(t) = M\psi(t) = Me^{-iHt}\psi(0)\,. \tag{4.11}$$

Since the two relations must hold for every initial state $\psi(0)$, we find the oeperator relation

$$e^{-iH't}M = Me^{-iHt}\,. \tag{4.12}$$

Using the (anti)unitarity of M, this can be written also as

$$e^{(-iH')t} = e^{M(-iH)M^\dagger t}\,. \tag{4.13}$$

Since this must be true at all times, we conclude[47]

$$-iH' = M(-iH)M^\dagger \qquad H' = (iM)H(iM)^\dagger \tag{4.14}$$

If the two observer are equivalent, then they must see the same dynamical laws and so the same Hamiltonian. In other words, for M to be a symmetry we require

$$H = (iM)H(iM)^\dagger\,. \tag{4.15}$$

For the most common case of unitary M, this means that

$$H = MHM^\dagger \Longrightarrow [M, H] = 0\,, \tag{4.16}$$

i.e., M commutes with the Hamiltonian. Conversely, an (anti)unitary M satisfying Eq. (4.15) leads to identical temporal evolutions for the two observers, and so to the same physics in the two corresponding reference frames.

Summarising, the dynamical requirement for a symmetry, i.e., that the transformed of the evolved is equal to the evolved of the transformed, reads then for a time-independent transformation

$$MU(t)\psi(0) = U(t)M\psi(0)\,. \tag{4.17}$$

The left-hand side can always be recast as

$$MU(t)\psi(0) = MU(t)M^\dagger M\psi(0) = U'(t)M\psi(0)\,, \tag{4.18}$$

which defines the temporal-evolution operator $U'(t)$. Imposing Eq. (4.17) implies that $U'(t) = U(t)$, and so that the corresponding Hamiltonians are identical (up to a sign in case of antiunitary M). Moreover, while the state observed at time t by $\mathcal{O}'$ differs in general from that observed by $\mathcal{O}$, the difference is due entirely to the fact that they observe a different initial state. If $\mathcal{O}$ and $\mathcal{O}'$ were to do separately the same experiment, i.e., each starting from the state corresponding for to the same vector $\psi(0)$ according to their own description, then they would observe the same vector $\psi(t)$ at later times, therefore seeing the same temporal evolution, and infering the same laws of physics.

[47] Notice that this implies that if H is time-independent, then so will be $H' = iM(-iH)M^\dagger$.

Group structure There is an important structure underlying the set of symmetry transformations of a physical system. If M_1 and M_2 are time-independent symmetry transformations, so is their composition $M = M_2 M_1$. Indeed, the product of unitary and/or antiunitary operators is still either unitary or antiunitary, and moreover

$$\begin{aligned}(iM)H(iM)^\dagger &= (iM_2M_1)H(iM_2M_1)^\dagger = [-i(iM_2iM_1)]H[-i(iM_2iM_1)]^\dagger \\ &= -i(iM_2)(iM_1)H(iM_1)^\dagger(iM_2)^\dagger i = -i(iM_2)H(iM_2)^\dagger i = -iHi = H\,.\end{aligned} \tag{4.19}$$

Such a composition is associative, i.e., if M_3 is another symmetry transformation then one has $M_3M_2M_1 = M_3(M_2M_1) = (M_3M_2)M_1$. The identity transformation is obviously a symmetry, and symmetry transformations are invertible (they are unitary or antiunitary transformations in the Hilbert space, but they were already invertible at the level of rays). The symmetry transformations of a physical system form therefore a *group*.

Time-dependent transformations The discussion above can be extended to time-dependent transformations $M(t)$ that do not alter time with a little effort. The request of invariance reads now

$$M(t)U(t)\psi(0) = U(t)M(0)\psi(0)\,, \tag{4.20}$$

expressing again that the transformed of the evolved is equal to the evolved of the transformed. Since this has to hold for any initial state, it follows that

$$M(t)U(t) = U(t)M(0) \Rightarrow M(t) = U(t)M(0)U(t)^\dagger\,. \tag{4.21}$$

Why is Eq. (4.20) the correct request? In general we can write

$$M(t)U(t)\psi(0) = M(t)U(t)M(0)^\dagger M(0)\psi(0) = \mathcal{U}(t)M(0)\psi(0)\,, \tag{4.22}$$

where $\mathcal{U}(t)$ provides the unitary temporal evolution of the system as seen by the observer $\mathcal{O}'$. We can obtain $\mathcal{U}(t)$ explicitly by writing down the differential equation it obeys, and imposing the initial condition $\mathcal{U}(0) = \mathbf{1}$. We have

$$\begin{aligned}\frac{d\mathcal{U}(t)}{dt} &= \frac{dM(t)}{dt}U(t)M(0)^\dagger + M(t)(-iHU(t))M(0)^\dagger \\ &= -i\left(M(t)HM(t)^\dagger + i\frac{dM(t)}{dt}M(t)^\dagger\right)\mathcal{U}(t) \equiv -i\mathcal{H}_M(t)\mathcal{U}(t)\,,\end{aligned} \tag{4.23}$$

where the time-dependent Hermitian operator $\mathcal{H}_M(t)$ is the Hamiltonian for $\mathcal{O}'$ (given that H is the Hamiltonian for $\mathcal{O}$). Hermiticity of $\mathcal{H}_M(t)$ follows from the fact that $M(t)$ is norm-preserving and onto:

$$0 = \frac{d}{dt}\mathbf{1} = \frac{d}{dt}[M(t)M(t)^\dagger] = \frac{d}{dt}[M(t)]M(t)^\dagger + M(t)\frac{d}{dt}[M(t)]^\dagger\,, \tag{4.24}$$

so that $[i(dM(t)/dt)M(t)^\dagger]^\dagger = -iM(t)(dM(t)^\dagger/dt) = i(dM(t)/dt)M(t)^\dagger$. The solution of Eq. (4.23) is the time-ordered exponential of $\mathcal{H}_M(t)$,

$$\begin{aligned}\mathcal{U}(t) &= \mathrm{Texp}\left\{-i\int_0^t dt'\,\mathcal{H}_M(t')\right\} \\ &\equiv \sum_{n=0}^{\infty}(-i)^n\int_0^t dt'_1\int_0^{t'_1} dt'_2\ldots\int_0^{t'_{n-1}} dt'_n\mathcal{H}_M(t'_1)\mathcal{H}_M(t'_2)\ldots\mathcal{H}_M(t'_n) \\ &= \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_0^t dt'_1\int_0^t dt'_2\ldots\int_0^t dt'_n\mathrm{T}\left(\mathcal{H}_M(t'_1)\mathcal{H}_M(t'_2)\ldots\mathcal{H}_M(t'_n)\right)\,,\end{aligned} \tag{4.25}$$

where the time-ordered product of operators is defined by ordering the operators according to decreasing time from left to right,

$$\mathrm{T}\left(A(t_1)\dots A(t_n)\right) = \theta(t_1-t_2)\dots\theta(t_{n-1}-t_n)A(t_1)\dots A(t_n) + \text{permutations}\,. \tag{4.26}$$

Clearly, if $\mathcal{H}_M(t) = H$ then $\mathcal{U}(t) = U(t)$, as they solve the same ordinary differential equation with the same initial condition; conversely, if $\mathcal{U}(t) = U(t)$ for all times then also their derivatives are equal, and so $\mathcal{H}_M(t) = H$. The two temporal evolutions will be the same if and only if the Hamiltonians $\mathcal{H}_M(t)$ and H are the same. If this is the case, then

$$H = M(t)HM(t)^\dagger + i\frac{dM(t)}{dt}M(t)^\dagger \Rightarrow -i[H, M(t)] = \frac{dM(t)}{dt}\,, \tag{4.27}$$

and this is solved by

$$M(t) = e^{-iHt}M(0)e^{iHt} = U(t)M(0)U(t)^\dagger \Rightarrow M(t)U(t) = U(t)M(0)\,. \tag{4.28}$$

Therefore, saying that the temporal evolution is the same for both observers is equivalent to saying that they use the same Hamiltonian to describe the system, and in turn this is equivalent to state that the evolved of the transformed is the transformed of the evolved. Eq. (4.28) also tells us that the dependence of the transformation $M(t)$ on time has to be entirely determined by the Hamiltonian of the system, with no room for some extra explicit dependence.

The group structure of symmetries stll holds if these time-dependent transformations are included. If M_1 and M_2 are two (now possibly time dependent) symmetry transformations, the product $M = M_2M_1$ is still unitary or antiunitary, and still satisfies Eq. (4.28),

$$M(t) = M_2(t)M_1(t) = U(t)M_2(0)M_1(0)U(t)^\dagger = U(t)M(0)U(t)^\dagger\,. \tag{4.29}$$

The other properties of a group are trivially verified.

Continuous and discrete symmetries In general, symmetries are classified in two big classes, namely *continuous* and *discrete* symmetries. A symmetry is continuous if it consists of a family of symmetry transformations parameterised by a set of continuous variables; it is discrete if there is no such family, i.e., if the transformations are "isolated". For continuous symmetries the representatives of the transformations on the Hilbert space of the system, $M = M(\alpha)$, depend on continuous parameters α. If they are connected to the identity, i.e., $M(\alpha_0) = \mathbf{1}$ for some α_0, then one can show that $M(\alpha)$ must be unitary [8]. Instead, discrete symmetries have to be discussed on a case-by-case basis.

One-parameter families of continuous transformations Let us focus on a continuous family of symmetry transformations $M(\alpha)$ characterised by a single real parameter α, choosing the parameterisation so that $M(0) = \mathbf{1}$. As we said above, $M(\alpha)$ are unitary operators, and being symmetries means that

$$[M(\alpha), H] = 0\,. \tag{4.30}$$

Taking the derivative with respect to α and then setting $\alpha = 0$ we find

$$[Q, H] = 0\,, \qquad Q \equiv -i\frac{dM}{d\alpha}\bigg|_{\alpha=0}\,. \tag{4.31}$$

The operator Q commutes with the Hamiltonian, and moreover it is Hermitean,

$$Q^\dagger = i\frac{dM^\dagger}{d\alpha}\bigg|_{\alpha=0} = -i\left(M^\dagger\frac{dM}{d\alpha}M^\dagger\right)\bigg|_{\alpha=0} = -i\frac{dM}{d\alpha}\bigg|_{\alpha=0} = Q\,, \tag{4.32}$$

having used the fact that $dU^\dagger = -U^\dagger dU U^\dagger$ for unitary operators. The operator Q then corresponds to an observable that is conserved in physical processes. To see this explicitly, recall that in the Heisenberg picture the observables are given a time dependence according to $\mathcal{O}(t) = e^{iHt}\mathcal{O}e^{-iHt}$. Then

$$\dot{Q}(t) = \frac{d}{dt}e^{iHt}Qe^{-iHt} = ie^{iHt}[H,Q]e^{-iHt} = i[H,Q(t)] = 0\,. \tag{4.33}$$

Any matrix element of $Q(t)$ is then clearly time-independent in the Heisenberg picture, and so in any picture.

There is actually more that we can do for one-parameter groups. For these groups one can choose the parameterisation so that,[48]

$$M(\alpha_1)M(\alpha_2) = M(\alpha_1+\alpha_2)\,, \tag{4.34}$$

again with $M(0) = \mathbf{1}$. Setting $\alpha_2 = \alpha$ and making $\alpha_1 \to d\alpha$ infinitesimal, and expanding in $d\alpha$, we find to lowest order

$$\begin{aligned} M(d\alpha)M(\alpha) &= M(\alpha+d\alpha)\,,\\ \left(M(0) + d\alpha\frac{dM}{d\alpha}(0)\right)M(\alpha) &= M(\alpha) + d\alpha\frac{dM}{d\alpha}(\alpha)\,,\\ \frac{dM}{d\alpha}(0)M(\alpha) &= \frac{dM}{d\alpha}(\alpha)\,. \end{aligned} \tag{4.35}$$

We recognise iQ on the left hand side of this differential equation,[49] which is easily solved to give

$$M(\alpha) = \exp\left\{\alpha\frac{dM}{d\alpha}(0)\right\} = \exp\{i\alpha Q\}\ . \tag{4.36}$$

As we showed above, since $M(\alpha)$ is a unitary symmetry one finds that Q is a self-adjoint operator commuting with H, and so a conserved physical quantity that can be diagonalised simultaneously with the Hamiltonian. Examples are energy and momentum, associated with the symmetry under temporal and spatial translations, and angular momentum, associated with the symmetry under rotations.

The construction above obviously does not apply to discrete symmetries, but there are nevertheless conservation laws associated with them. Discrete symmetries include most notably parity (P), charge conjugation (C), and time reversal (T). In the following subsection we discuss them in some detail.

[48] For a one-parameter group we require in general that $M(\alpha_1)M(\alpha_2) = M(f(\alpha_1,\alpha_2))$ for some function f. Under reasonable smoothness conditions, one can always choose the parameterisation so that Eq. (4.34) holds, and that $\alpha = 0$ corresponds to the identity. See, e.g., Ref. [9].

[49] The definitions of Q with the current parameterisation can at most differ from the previous one by an irrelevant constant factor,

$$Q_{\rm old} = \frac{dM(\alpha_{\rm old})}{d\alpha_{\rm old}}\Big|_{\alpha_{\rm old}=0} = \frac{dM(\alpha_{\rm new})}{d\alpha_{\rm new}}\Big|_{\alpha_{\rm new}(0)=0}\frac{d\alpha_{\rm new}(\alpha_{\rm old})}{d\alpha_{\rm old}}\Big|_{\alpha_{\rm old}=0} = Q_{\rm new}\frac{d\alpha_{\rm new}(\alpha_{\rm old})}{d\alpha_{\rm old}}\Big|_{\alpha_{\rm old}=0}\,.$$

4.1 Discrete symmetries

We have discussed above the issue of symmetry on general grounds, without specifying a physical system. We now specialise to the case of free particles, i.e., localised objects travelling on straight lines at constant speed. This choice is made both because of its simplicity, and because of its practical relevance: the typical high-energy experiment consists in taking two bunches of particles and shooting them at each other. In the initial stages of the experiments these particles are far away from each other, so not interacting yet, and when measurements are done on the final products of the process these are again far away from each other, so not interacting any more. As we have already discussed, to a very good approximations these experiments involve then free particles both in the initial and in the final state.

Free particle states A free particle is characterised by its type, as defined by its mass m, spin s, electric charge q and possibly other conserved (and compatible) charges, like, e.g., baryon number and lepton family number, plus its energy E, momenta $\vec{p}$ and the component of the spin in some prescribed direction, conventionally taken to be the third, or z, component, i.e., s_z. Once that the particle type is specified, $\vec{p}$ and s_z constitute a complete set of compatible observables, with energy determined from the momenta (and the mass) through the dispersion relation $E^2 = \vec{p}^2 + m^2$. When writing the state vector of a free particle, we usually put mass and the other defining observables under the label of the particle type; if the state is an eigenstate of momenta and s_z, we further specify their values. In Dirac notation, we would write, e.g., for a momentum eigenstate of a neutral pion $|\vec{p}; \pi^0\rangle$ (there is no spin here); for a proton, $|\vec{p}, s_z; p\rangle$; and so on. Other notations might be used, depending on what one wants to emphasise. One may choose to use a different complete set of observable, e.g., replacing $\vec{p} \to E, \ell, \ell_z$, where ℓ is the total orbital angular momentum and ℓ_z its component in the z-direction; in this case one would write for a pion $|E, \ell, \ell_z; \pi^0\rangle$. Of course, any superposition of these basis states is allowed.

We now discuss the effect of the discrete symmetries P, C, and T on the free particle states.

4.1.1 Parity

Parity (P) consists in the change of the sign of all the spatial coordinates of our reference frame. In non-relativistic quantum mechanics, the effect of this transformation on the state of a particle is simply to change its wave function to $P\psi_{s_z}(\vec{x}) = \psi_{s_z}(-\vec{x})$, with spin being unaffected. In the relativistic case, however, we cannot use wave functions in coordinate space to describe our system. Nonetheless, we can define the parity transformation on states by observing that under the required change of coordinates, all the components of the momentum of a particle will change sign, while angular momenta (and spin in particular) will remain unchanged. We must then have for the state of a particle of type α with momentum $\vec{p}$ and z-component of the spin s_z,

$$P|\vec{p}, s_z; \alpha\rangle = \eta_\alpha |-\vec{p}, s_z; \alpha\rangle \,, \tag{4.37}$$

where P denotes the unitary or antiunitary operator implementing the parity transformation on the Hilbert space of the particle. The quantity η_α is a phase factor named *intrinsic parity*, which does not change the physical content of a parity transformation, but that has to be included for generality:[50] if P is a symmetry, then a consistent assignment of phases in Eq. (4.37) can be

[50] Dependence of the intrinsic parity on anything other than the particle type α is excluded by symmetry under rotations, that commute with P.

made.

One can equivalently discuss parity in a Heisenberg-type picture, in which the quantum states are given once and for all, and for all observers, with the effect of symmetry transformations shifted on the operators. In this picture one finds for the momentum and angular momentum *operators*

$$P^\dagger \vec{p} P = -\vec{p}, \qquad P^\dagger \vec{J} P = \vec{J}, \tag{4.38}$$

which since the spin of a particle is its angular momentum in its rest frame, is equivalent to the transformation rule Eq. (4.37).

Unitarity of P We have not decided yet if parity has to be realised in the Hilbert space of particles as a unitary or an antiunitary transformation. We now argue that the unitary option must be chosen. The reason for this is physical, and comes from the requirement of invariance that, since P does not affect time, reads $PU(t) = U(t)P$. For infinitesimal t we then find $PiH = iHP$. For linear unitary P we find $[P, H] = 0$, but for antilinear antiunitary we would have instead $PH + HP = \{P, H\} = 0$, so that to every state ψ_E with energy E it would correspond a state $\psi_E^{(P)} = P\psi_E$ with energy $H\psi_E^{(P)} = HP\psi_E = -PH\psi_E = -E\psi_E^{(P)}$. Since negative energy particle states are not found in nature, we are forced to choose P to be unitary and commuting with the Hamiltonian. This means in particular that H and P can be diagonalised simultaneously. In this case, instead of momentum eigenstates one would use energy and orbital angular momentum eigenstates $|E, \ell, \ell_z; \alpha\rangle$, that obey $P|E, \ell, \ell_z; \alpha\rangle = (-1)^\ell |E, \ell, \ell_z; \alpha\rangle$, in full analogy with the non-relativistic result $\psi_{\ell\ell_z}(-\vec{x}) = (-1)^\ell \psi_{\ell\ell_z}(\vec{x})$.

Non-uniqueness of intrinsic parities The assignment of intrinsic parities is in general not unique: if there is a continuous group of phase transformations generated by some operator Φ that is a symmetry of the system, then we can redefine parity to be $P' = Pe^{i\Phi}$: this is still a symmetry which does what parity has to do on physical states. If we limit ourselves to a world in which only strong and electromagnetic interactions are present, and the only matter particles are the up quark, the down quark, the electron, and their antiparticles, then there are three such generators, namely the electric charge Q, the baryon number B and the lepton number L, so that we can fix the phases of, say, the proton, the neutron and the electron to 1. If $P^{(0)}$ is the initial definition of the parity operator, with corresponding intrinsic parities $\eta_\alpha^{(0)}$, then setting $P = P^{(0)} e^{i(\alpha B + \beta L + \gamma Q)}$ we can choose α, β and γ such that

$$\begin{aligned} \text{proton:} \qquad & \eta_p = \eta_p^{(0)} e^{i(\alpha+\gamma)} = 1\,, \\ \text{neutron:} \qquad & \eta_n = \eta_n^{(0)} e^{i\alpha} = 1\,, \\ \text{electron:} \qquad & \eta_e = \eta_e^{(0)} e^{i(\beta-\gamma)} = 1\,. \end{aligned} \tag{4.39}$$

All the other intrinsic parities are now fixed by consistency. In general, we can choose arbitrarily one intrinsic parity for each conserved quantity: if we add, say, the muon to our particle zoo, then we can fix its intrinsic parity to 1 using the muonic lepton number. Instead, for truly neutral particles like, e.g., the photon or the neutral pion, the intrinsic parity cannot be redefined through a phase transformation, and therefore carries a genuine intrinsic meaning.

Intrinsic parities of antiparticles From its definition, the parity operator is such that P^2 is just a phase transformation of each state. If this transformation belongs to a continuous set

of phase transformation symmetries like the ones described above, then it is possible to redefine it such that $P^2 = \mathbf{1}$. This is the case in the Standard Model, and so we can take without loss of generality $\eta_\alpha^2 = 1$, i.e., $\eta_\alpha = \pm 1$. The reason is the existence of a sufficient number of conserved charges, and the absence of self-conjugate fermions. In fact, quantum field theory imposes that the intrinsic parities of a particle α and its antiparticle $\bar{\alpha}$ be related as

$$\eta_\alpha \eta_{\bar{\alpha}} = \begin{cases} +1 & \text{if they are bosons,} \\ -1 & \text{if they are fermions.} \end{cases} \tag{4.40}$$

For self-conjugate bosons, whose intrinsic phase *cannot* be redefined by a phase transformation, this does not contradict $P^2 = 1$. For a self-conjugate fermion (a *Majorana fermion*), one would find instead $\eta_\alpha = \pm i$, so that $P^2 \neq 1$. However, such particles have never been observed so far.

Determination of intrinsic parities It is possible to assign intrinsic parities to particles empirically, using conservation of parity in a physical process, i.e., that parity is a symmetry (of course, when this applies) and the conventionally chosen values. Consider scattering or decay processes of the form

$$a\ b \to c\ d\,, \qquad\qquad a \to b\ c\,. \tag{4.41}$$

We know that the transition probability for scattering processes are obtained from the matrix elements of the S operator [see Eqs. (2.26) and (2.52)]. For decay processes a similar operator, the decay operator Γ, is used instead. We have

$$\text{transition probability } \propto\ |\langle c\ d|S|a\ b\rangle|^2\,, \quad |\langle b\ c|\Gamma|a\rangle|^2\,. \tag{4.42}$$

The very fact that a process happens implies that these matrix elements are nonzero. Consider now states with well-defined energy and orbital angular momentum, instead of momentum eigenstates, that as we said above obey $P|\ell\ell_z\rangle = (-1)^\ell|\ell\ell_z\rangle$. If P is a symmetry of H, then since it is a symmetry of H_0 we also have that $[P, S] = [P, \Gamma] = 0$, and so (omitting all irrelevant quantities from the notation)

$$\begin{aligned} 0 &= \langle \ell', \ell'_z; c\ d|[P, S]|\ell, \ell_z; a\ b\rangle = [(-1)^{\ell'}\eta_c\eta_d - (-1)^\ell\eta_a\eta_b]\langle \ell', \ell'_z; c\ d|S|\ell, \ell_z; a\ b\rangle\,, \\ 0 &= \langle \ell, \ell_z; b\ c|[P, \Gamma]|a\rangle = [(-1)^\ell\eta_b\eta_c - \eta_a]\langle \ell, \ell_z; b\ c|\Gamma|a\rangle\,, \end{aligned} \tag{4.43}$$

where in the case of a decay process we work in the rest frame of the decaying particles, so that there is no orbital angular momentum. Since the matrix elements of S and Γ are nonvanishing, Eq. (4.43) yields the relations

$$(-1)^{\ell'}\eta_c\eta_d = (-1)^\ell\eta_a\eta_b\,, \qquad (-1)^\ell\eta_b\eta_c = \eta_a\,. \tag{4.44}$$

Using the conventional intrinsic parities and those that have already been determined, we can go on and assign an intrinsic parity to one of the particles involved in the process.

To see how things work in practice, let us consider an example, that of the parity of the charged pion. The relevant physical process is that of pion capture by a deuteron (d), which is a bound state of a proton and a neutron, $d = (pn)$, with orbital angular momentum $\ell_d = 0$. The intrinsic parity of the deuteron is easily determined as $\eta_d = \eta_p\eta_n(-1)^0 = 1$. Finally, the spin of the deuteron is $s_d = 1$. The relevant physical process is

$$\pi^-\ d \to n\ n\,, \tag{4.45}$$

ℓ	S	J	$(-1)^{S+\ell+1}$
0	1	1	1
1	0	1	1
1	1	$0 \oplus 1 \oplus 2$	-1
2	1	$1 \oplus 2 \oplus 3$	1

Table 7: Combinations of S and ℓ allowed by the conservation of angular momentum in the $\pi^- d$ capture process.

which proceeds through the formation of a $\pi^- d$ atom and its subsequent decay into a pair of neutrons. This decay takes place from the ground state of the pion-deuteron atom, which has $\ell = 0$ (with a small admixture of $\ell = 2$): its phase under a parity transformations is $(-1)^0 = 1$. Conservation of angular momentum implies, since $\ell = 0$ and the pion is spinless, that $J = 1$ both in the initial and in the final state. The final state is nonrelativistic, and can therefore be described in the framework of quantum mechanics as $\psi(\vec{x})|S, S_z\rangle$, with $\psi(\vec{x}) = R_\ell(r)Y_\ell^m(\theta, \varphi)$ the spatial wave function and $|S, S_z\rangle$ the spin wave function. Overall the wave function has to be antisymmetric under exchange of the neutrons since these are fermions. The spin wave function is constructed starting from two spin-$\frac{1}{2}$ states, and since $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$, we have $S = 0, 1$. It is easy to write down explicitly the states corresponding to the two cases,

$$\begin{aligned} |\tfrac{1}{2}\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}\tfrac{1}{2}\rangle &= |11\rangle \\ \tfrac{1}{\sqrt{2}}\left(|\tfrac{1}{2}-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}\tfrac{1}{2}\rangle + |\tfrac{1}{2}\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}-\tfrac{1}{2}\rangle\right) &= |10\rangle \\ |\tfrac{1}{2}-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}-\tfrac{1}{2}\rangle &= |1-1\rangle \\ \tfrac{1}{\sqrt{2}}\left(|\tfrac{1}{2}-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}\tfrac{1}{2}\rangle - |\tfrac{1}{2}\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}-\tfrac{1}{2}\rangle\right) &= |00\rangle\,, \end{aligned} \tag{4.46}$$

and in turn to determine the sign acquired under exchange of the two neutrons as $(-1)^{S+1}$. As for the spatial part, exchanging the neutrons corresponds to sending $\vec{x} \to -\vec{x}$, so the corresponding sign is $(-1)^\ell$. All in all we must have $(-1)^{S+\ell+1} = -1$. Conservation of angular momentum limits the possible combinations of S and ℓ in the final state to those listed in Table 7, since those are the only ones containing 1. Among these, the only one allowed by the Fermi-Dirac statistics of neutrons is $\ell = S = 1$. Since $\eta_n = 1$, this implies for the parity of the final state $\eta_n^2(-1)^\ell = -1$. This must be equal to that in the initial state, which is $\eta_\pi \eta_d (-1)^0 = \eta_\pi \eta_d = \eta_\pi$. In conclusion then $\eta_\pi = -1$.

Another example is the intrinsic parity of the Δ^{++}, determined from the decay $\Delta^{++} \to p\,\pi^+$. Since $s_{\Delta^{++}} = \frac{3}{2}$, $s_p = \frac{1}{2}$ and $s_{\pi^+} = 0$, the final state must have either $\ell = 1$ or $\ell = 2$. This can be determined from the angular distribution of the decay products. From the relation $\eta_{\Delta^{++}} = \eta_p \eta_{\pi^+}(-1)^\ell = (-1)^{\ell+1}$, and the experimental determination that $\ell = 1$, one finds $\eta_{\Delta^{++}} = 1$.

In other cases, the intrinsic parity can be assigned on the basis of theoretical considerations. This is the case of the photon: classically the electric field $\vec{E}$ transforms like a vector, and since $\vec{E} = -\vec{\nabla}\phi - \frac{\partial}{\partial t}\vec{A}$, so has to transform the vector potential $\vec{A}$. Vectors change sign under parity. Upon quantisation the physical modes of the photon are encoded in $\vec{A}$, and for the intrinsic parity this leads to $\eta_\gamma = -1$. Another way to see this is that in the quantum theory of electrodynamics (QED), the coupling of photons to electrons is described by means of the photon field A_μ and of the electric current J^μ. The electric current is a Lorentz vector, and so has to be the photon, hence $\eta_\gamma = -1$.

4.1.2 Charge conjugation

We now turn to charge conjugation (C), which consists in exchanging particles with the corresponding antiparticles, keeping momenta and spin unchanged. Denoting with $\bar{\alpha}$ the antiparticle corresponding to particle α, the action of C on particle states is defined by

$$C|\vec{p}, s_z; \alpha\rangle = \xi_\alpha |\vec{p}, s_z; \bar{\alpha}\rangle \,, \tag{4.47}$$

where again a phase ξ_α has to be included for generality.[51] An argument identical to that used above for parity requires that C be a unitary operator and $[C, H] = 0$.

As we have already said, changing from particle to antiparticle leads to changing sign of to all the charges, like, e.g., electric charge (q), baryon (B), lepton (L) and lepton family (L_ℓ) numbers. Moreover, since the magnetic moment of a particle, $\vec{\mu}$, is proportional to the product of charge and spin, $\vec{\mu} \propto q\vec{s}$, it also changes sign under C. For all these quantities we have that $\{C, \mathcal{O}\} = 0$. If we reversed the discussion and started from asking that an operator C exists obeying this anticommutation relation, then we would find that for each particle state with given values of these observables there is a corresponding state for which the observables have the same magnitude but opposite sign: indeed, these are the antiparticles.

Intrinsic charge-conjugation phase of antiparticles Applying C twice we find

$$C^2|\vec{p}, s_z; \alpha\rangle = \xi_\alpha \xi_{\bar{\alpha}} |\vec{p}, s_z; \alpha\rangle \,, \tag{4.48}$$

i.e., C^2 is just a phase transformation. Quantum field theory requires that $\xi_\alpha \xi_{\bar{\alpha}} = 1$, both for bosons and fermions, and so we can write $C^2 = \mathbf{1}$. In general, however, the value of $\xi_\alpha = \pm 1$ is relevant only for neutral particles that are self-conjugate, like γ and π^0 (but not n), while it can be adjusted freely for non-self-conjugate particles.

Determination of intrinsic charge-conjugation phases In order to assign the intrinsic charge conjugation phase ξ_α to a self-conjugate particle we can either rely on theoretical arguments or on the selection rule implied by charge conjugation invariance of a theory. As an example of the first method, consider the photon. Recall the classical Maxwell equations with sources, and the relations between the electric and magnetic fields and the potential,

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} \propto \rho \,, \qquad & \vec{E} = -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \,, \\ \vec{\nabla} \wedge \vec{B} \propto \vec{J} \,, \qquad & \vec{B} = \vec{\nabla} \wedge \vec{A} \,. \end{aligned} \tag{4.49}$$

Exchanging negative and positive charges corresponds to $\rho \to -\rho$ and $\vec{J} \to -\vec{J}$, which in turn changes the signs of $\vec{E}$ and $\vec{B}$. At the level of the potential $A_\mu = (\phi, \vec{A})$, this is obtained by changing $A_\mu \to -A_\mu$. Carrying this over to the quantum case we then have for the quantum photon field $C^\dagger A_\mu C = -A_\mu$, and so $\xi_\gamma = -1$.

This result can then be employed to establish the charge conjugation phase of the neutral pion by means of the second method. Since the π^0 decays into two photons, we must have $\xi_{\pi^0} = \xi_\gamma^2 = 1$. The same value of ξ can be assigned to the charged pions if we wish, but this is

[51] Like with parity, ξ_α depends only on α due to symmetry under rotations, since these commute with C.

just a matter of convention: since they are not self-conjugate, there is no selection rule for them associated with charge conjugation, and we cannot fix $\xi_{\pi^\pm}$ in this way.

If charge conjugation were an exact symmetry of nature, then the decay process $\pi^0 \to \gamma\gamma\gamma$ would be strictly forbidden. Violations of charge conjugation symmetry come only from the weak interactions, which have very little to do with this process, so we expect that this process is strongly suppressed. Experimental results give for the relative probability of $\pi^0 \to 3\gamma$ with respect to $\pi^0 \to 2\gamma$ the upper bound $\Gamma_{\pi^0\to 3\gamma}/\Gamma_{\pi^0\to 2\gamma} < 3.1 \cdot 10^{-8}$.

4.1.3 Time reversal and CPT

We conclude with a brief discussion of time reversal (T), i.e., the inversion of the arrow of time. Under such a transformation, the signs of both the momentum components and the spin components change sign. The effect of T on particle states reads

$$T|\vec{p}, s_z; \alpha\rangle = \zeta_{\alpha,s_z}|-\vec{p}, -s_z; \alpha\rangle\,, \tag{4.50}$$

where the extra phase factor this time depends on s_z as well as on the particle species, and reads $\zeta_{\alpha,s_z} = (-1)^{s-s_z}\zeta_\alpha$.

Antiunitarity of T Contrary to P and C, T is an antiunitary symmetry. In fact, in contrast with Eq. (4.17), the requirement of invariance reads here $TU(t)\psi(0) = U(-t)T\psi(0)$: in fact, assuming the same time origin for both observers, the times at which the system is observed are 0 and t in the first frame, and 0 and $-t$ in the second frame, since in the latter time flows in the opposite direction. Since $\psi(0)$ and t are arbitrary, one finds

$$TU(t) = U(-t)T \Rightarrow TiH = -iHT\,. \tag{4.51}$$

There are two alternatives: if T is antilinear and antiunitary, then $Ti = -iT$ and we need $[T, H] = 0$, while if T is linear and unitary then we need $\{T, H\} = 0$. The second case is excluded again by the experimental absence of negative energy particle states, which forces us to have T antiunitary and commuting with the Hamiltonian. As a consequence of antiunitarity, the residual phase ζ_α has no physical meaning, since it can be reabsorbed in a redefinition of the particle states: in fact, if we redefine $|\vec{p}, s_z; \alpha\rangle \to e^{i\phi}|\vec{p}, s_z; \alpha\rangle$, then

$$Te^{i\phi}|\vec{p}, s_z; \alpha\rangle = e^{-i\phi}T|\vec{p}, s_z; \alpha\rangle = e^{-i\phi}\zeta_\alpha(-1)^{s-s_z}|-\vec{p}, -s_z; \alpha\rangle\,, \tag{4.52}$$

and we are free to choose $e^{-i\phi}\zeta_\alpha = 1$.

The CPT theorem It is a general theorem of quantum field theory, the *CPT theorem*, that for any Lorentz-invariant theory of local quantum fields, the antiunitary transformation $\Theta = CPT$ is a symmetry. On a particle state Θ acts as follows:

$$\begin{aligned}\Theta|\vec{p}, s_z; \alpha\rangle &= CPT|\vec{p}, s_z; \alpha\rangle = CP\zeta_{\alpha,s_z}|-\vec{p}, -s_z; \alpha\rangle = C\eta_\alpha\zeta_{\alpha,s_z}|\vec{p}, -s_z; \alpha\rangle \\ &= \xi_\alpha\eta_\alpha\zeta_{\alpha,s_z}|\vec{p}, -s_z; \bar{\alpha}\rangle = \theta_{\alpha,s_z}|\vec{p}, -s_z; \bar{\alpha}\rangle\,,\end{aligned} \tag{4.53}$$

Despite the nonconservation of P and C (and also CP) in weak interactions, the product CPT is a good symmetry also in that case. If we were to find violations of this symmetry, this would

have the deeply troubling consequence that quantum field theory could not be employed to explain them, and should be replaced as the framework of our fundamental theories of pyhsics.

The CPT theorem has interesting consequences. As an example, it implies that particles and antiparticles must have the same mass. For unstable particles, it implies that the lifetime is the same as that of the corresponding antiparticle. The proof of these statements is rather straightforward. For the equality of masses we have

$$\begin{aligned}\Theta p^2|\vec{p},s_z;\alpha\rangle &= m_\alpha^2\theta_{\alpha,s_z}|\vec{p},-s_z;\bar{\alpha}\rangle \\ &= \Theta p^2\Theta^\dagger\Theta|\vec{p},s_z;\alpha\rangle = p^2\theta_{\alpha,s_z}|\vec{p},-s_z;\bar{\alpha}\rangle = m_{\bar{\alpha}}^2\theta_{\alpha,s_z}|\vec{p},-s_z;\bar{\alpha}\rangle\,,\end{aligned} \tag{4.54}$$

since the four-momentum operator $p=(p^0=H,\vec{p})$ transforms as $TpT^\dagger = PpP^\dagger = (p^0,-\vec{p})$, and $CpC^\dagger = p$. It then follows $m_\alpha = m_{\bar{\alpha}}$. One can similarly prove that for the spin one has $s_\alpha = s_{\bar{\alpha}}$. For the lifetimes of unstable particles, working in the Born approximation for decay probabilities, we have $\tau_\alpha^{-1} = \Gamma_\alpha = \sum_f c_f|\langle f|H_I|\alpha\rangle|^2$, where c_f are kinematical factors that depend on the mass and spin of the final state particles (as well as their momenta, which are however summed over). We find

$$\begin{aligned}\Gamma_\alpha &= \sum_f c_f|\langle f|H_I|\alpha\rangle|^2 = \sum_f c_f|\langle f|\Theta^\dagger H_I\Theta|\alpha\rangle|^2 \\ &= \sum_f c_f|\langle \bar{f}|H_I|\bar{\alpha}\rangle|^2 = \sum_f c_{\bar{f}}|\langle \bar{f}|H_I|\bar{\alpha}\rangle|^2 = \Gamma_{\bar{\alpha}}\,.\end{aligned} \tag{4.55}$$

4.2 Continuous symmetries: rotations

It is an experimental fact that a massive particle in its rest frame can be in more than one physical state, i.e., the "particle is at rest" state characterised by vanishing spatial momentum $\vec{p}=\vec{0}$. In general, the states of a particle at rest form a finite-dimensional Hilbert space. We denote such states as $|\vec{0},\psi\rangle \in \mathcal{H}_{\rm rest}$, with $\mathcal{H}_{\rm rest}$ finite dimensional, and decompose them on an orthonormal basis as $|\vec{0},\psi\rangle = \sum_n \psi_n|\vec{0},n\rangle$. For example, an electron at rest can be found in any superposition of the "spin-up" and "spin-down" states – whatever those are. A particle in motion with momentum $\vec{p}$ is related to a particle at rest by a suitable Lorentz boost $\Lambda_{\vec{p}}$, with $\Lambda(m,\vec{0}) = (E(\vec{p}),\vec{p})$. This transformation belongs to a continuous group of symmetries (the Lorentz group), and as such it is realised on the Hilbert space of states by a unitary operator $U(\Lambda_{\vec{p}})$. The most general state of particle will then be expressible as a linear superposition of states $|\vec{p},n\rangle \equiv U(\Lambda_{\vec{p}})|\vec{0},n\rangle$. We are left with the task of characterising the states $|\vec{0},n\rangle$. In this section we discuss what these states are, so explaining what "spin-up" and "spin-down" electron states are; and show that this states are related to the rotational properties of particles. These are encoded in a new quantum number called *spin*.

We notice first that the rest frame of a particle is not unique: any rotation of the coordinate system still yields a frame where the particle is at rest. Since rotations are symmetries, two observers whose coordinate systems are related by a rotation will assign vectors out of the same (finite-dimensional) Hilbert space to the states of a particle at rest. The two observers will in general assign different vectors to a given state of the particle; since rotations form a continuous group of symmetries, by Wigner's and Bargmann's theorems we can relate the vectors assigned to the states by the two observers by means of a unitary operator. If the "unrotated" observer describes the state of the particle with the vector $|\psi\rangle$, and the "rotated" observer whose

coordinate system is obtained with a rotation R from that of the unrotated observer describes it with the vector $|\psi_R\rangle$, then

$$|\psi_R\rangle = U(R)|\psi\rangle\,, \tag{4.56}$$

with $U(R)$ an R-dependent unitary operator. If $|n\rangle \equiv |\vec{0}, n\rangle$ are the basis states for our particle at rest, then $\psi\rangle = \sum_n \psi_n |n\rangle$ and

$$|\psi_R\rangle = U(R)|\psi\rangle = \sum_n \psi_n U(R)|n\rangle = \sum_{n,n'} \psi_n D_{n'n}(R)|n'\rangle\,, \tag{4.57}$$

where $U(R)|n\rangle = D_{n'n}(R)|n'\rangle$ is the transformed of $|n\rangle$, and $D_{n'n}(R)$ are the components of a square matrix $D(R)$,

$$D_{n'n}(R) = \langle n'|U(R)|n\rangle\,. \tag{4.58}$$

Since $U(R)$ is unitary,

$$\begin{aligned}\delta_{n'n} = \langle n'|n\rangle = \langle n'|U(R)^\dagger U(R)|n\rangle &= \sum_{\tilde{n}',\tilde{n}} \langle \tilde{n}'|\tilde{n}\rangle D_{\tilde{n}'n'}(R)^* D_{\tilde{n}n}(R) \\ &= \sum_{\tilde{n}',\tilde{n}} D_{\tilde{n}'n'}(R)^* D_{\tilde{n}n}(R)\delta_{\tilde{n}'\tilde{n}} = [D(R)^\dagger D(R)]_{n'n}\,,\end{aligned} \tag{4.59}$$

i.e., $D(R)$ are unitary matrices.

4.2.1 Group representations

Consistency of the description requires that if we change coordinates by a composite rotation R_2R_1, the resulting state should be the same independently of whether we treat R_2R_1 as a single rotation, or as the composition of R_1 followed by R_2. The mathematical translation of this statement is

$$U(R_2R_1)|\psi\rangle = U(R_2)U(R_1)|\psi\rangle\,, \tag{4.60}$$

that should hold for any vector, and so

$$U(R_2R_1) = U(R_2)U(R_1)\,, \tag{4.61}$$

or in terms of $D(R)$

$$D(R_2R_1) = D(R_2)D(R_1)\,. \tag{4.62}$$

As is well known, rotations form a group, namely the group $SO(3)$ of 3×3 orthogonal unimodular matrices, obeying $O^TO = \mathbf{1}$ and $\det O = 1$. A mapping from a group G to the space of invertible $n \times n$ complex matrices $\mathrm{GL}(\mathbb{C}, n)$, associating a matrix $D(g)$ to each element of G,

$$\begin{aligned} D:\ G &\to \mathrm{GL}(\mathbb{C}, n) \\ g &\to D(g)\,, \end{aligned} \tag{4.63}$$

and obeying the relation Eq. (4.62), i.e.,

$$D(g_1g_2) = D(g_1)D(g_2)\,, \tag{4.64}$$

is said to provide a *representation* of the group. In practice, a representation associates a matrix to each group element, and the composition law of the group is reproduced by the

composition law of these matrices. A representation in terms of unitary matrices is called a *unitary representation*. The acceptable ways to implement the effect of rotations on the state vectors of a physical system are then restricted to the unitary representations of the rotation group. Strictly speaking, physical equivalence requires only that Eq. (4.62) holds up to a physically irrelevant phase (which is experimentally unobservable). A mapping from a group to a space of invertible matrices obeying $D(g_1 g_2) = e^{i\phi(g_1,g_2)} D(g_1) D(g_2)$ is said to provide a *projective representation* of the group. More precisely, then, rotations must be implemented through a projective representation of the rotation group.

The basic relation Eq. (4.64) automatically implies the following results. For the neutral element e,

$$D(g) = D(eg) = D(e)D(g) \Longrightarrow D(e) = \mathbf{1}\,, \tag{4.65}$$

i.e., e is mapped to the identity matrix. Moreover, for the inverse element g^{-1} we find

$$\mathbf{1} = D(e) = D(gg^{-1}) = D(g)D(g^{-1}) \Longrightarrow D(g^{-1}) = [D(g)]^{-1}\,. \tag{4.66}$$

It is possible that all the matrices $D(g)$ leave some subspace $S \subseteq \mathbb{C}^n$ invariant, i.e., for every vector $s \in S$ one has that $s_g = D(g)s \in S$ is still in S. More compactly, we write $D(g)S \subseteq S$ $\forall g \in G$. Such a subspace is called and *invariant subspace*, and if there is at least one nontrivial such subspace (i.e., $S \neq \mathbb{C}^n$ and $S \neq \{0\}$, where 0 is the zero vector in $\mathbb{C}^n$), then we say that the representation is *reducible*. The corresponding matrices can then *all* be put in the form

$$D(g) = \begin{pmatrix} D_{11} & D_{12} & \dots & D_{1M} \\ 0 & D_{22} & \dots & D_{1M} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & D_{MM} \end{pmatrix} \tag{4.67}$$

by a suitable change of basis, with the block D_{jj} corresponding to the mapping of the invariant subspaces into themselves. If the basis can be chosen so that $D_{ij} = 0$ for $i \neq j$ for all g, then the whole of $\mathbb{C}^n$ splits up in several invariant subspaces, and the representation is said to be *completely reducible*. If instead no nontrivial invariant subspace exists, then the representation is *irreducible*. A completeley reducible representation can then be brought to block diagonal form, with each block corresponding to an invariant subspace not containing any smaller (nontrivial) invariant subspace, and providing an irreducible representation of the group,

$$D(g) = \begin{pmatrix} D_{11} & 0 & \dots & 0 \\ 0 & D_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & D_{MM} \end{pmatrix}\,. \tag{4.68}$$

For unitary representations, $D(g)^\dagger D(g) = \mathbf{1}$, and so $D(g^{-1}) = D(g)^{-1} = D(g)^\dagger$. Such representations are always completely reducible. In fact, if S is an invariant subspace, then its orthogonal complement $S^\perp$ is also invariant. By definition, $S^\perp = \{w \mid (v,w) = 0,\ \forall v \in S\}$, and so for every $v \in S$ and $w \in S^\perp$,

$$(v, D(g)w) = (D(g)^\dagger v, D(g)^\dagger D(g) w) = (D(g^{-1})v, w) = 0\,, \tag{4.69}$$

since $D(g^{-1})v \in S$ by definition of invariant subspace, and therefore $D(g)w \in S^\perp$. Repeating the argument until no more invariant subspaces are found one ends up with all the $D(g)$ in

the block-diagonal form of Eq. (4.68), with each block providing an irreducible representation of the group. Since we are interested in finding unitary representations of SO(3), our task is completed if we can find all the irreducible representations of SO(3), or more precisely its irreducible projective representations.

4.2.2 Lie groups and Lie algebras

The group of rotations, SO(3), belongs to a special type of groups known as *(matrix) Lie groups*. Lie groups are at the same time groups and smooth manifolds, i.e., their elements $g = g(\alpha)$ can be parameterised in terms of a set of continuous real variables $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, and form a continuous space that is locally similar to the Euclidean space $\mathbb{R}^d$.[52] The integer d is the called the dimension of the group. Groups that are (simply) connected as a manifold are called *(simply) connected Lie groups*. The component of a Lie group connected to the identity element can essentially be entirely reconstructed starting from the elements near the identity, usually associated with $\alpha = 0$ by a convenient choice of parameterisation. Near $\alpha = 0$,

$$g(\alpha) = g(0) + \alpha_a \frac{\partial g}{\partial \alpha_a}\bigg|_{\alpha=0} + \ldots = \mathbf{1} + i\alpha_a L_a + \ldots\,, \qquad L_a \equiv -i\frac{\partial g}{\partial \alpha_a}\bigg|_{\alpha=0}\,. \tag{4.70}$$

The matrices L_a, $a = 1, \ldots, d$, are the *generators* of the group. One can prove the following results:

- the matrices $g_0(\alpha) = e^{i\alpha\cdot L}$ obtained as the matrix exponential of $i\alpha \cdot L$, are elements of a neighbourhood of $\mathbf{1}$;
- the most general g in the Lie group component connected to the identity is obtained as a finite product of such g_0s.

Moreover, the generators obey closed commutation relations,

$$[L_a, L_b] = iC_{ab}{}^c L_c\,, \tag{4.71}$$

with real coefficients $C_{ab}{}^c$ called *structure constants*. The real linear space spanned by $\{L_a\}$, together with the commutation relations Eq. (4.71), define the *Lie algebra* associated with the Lie group. In general, different Lie groups can have the same Lie algebra.

The reason for this detour on Lie groups and Lie algebra is that representations of a Lie group can be reconstructed from those of its Lie algebra, which are easier to construct. A representation of a Lie algebra is a linear mapping $d(X) = X^a d(L_a)$ of its elements $X = X^a L_a$, $X^a \in \mathbb{R}$ into a space of matrices (non necessarily invertible this time), that respects the commutation relations Eq. (4.71), i.e.,

$$[d(L_a), d(L_b)] = iC_{ab}{}^c d(L_c) = d([L_a, L_b])\,. \tag{4.72}$$

Irreducible representations are again defined as those that do not leave any subspace invariant. Given a representation $D(g) = D(g(\alpha))$ of the group, a representation of its algebra can always be obtained by differentiation through $d(L_a) = -i\frac{\partial}{\partial\alpha_a} D(g(\alpha))|_{\alpha=0}$. Conversely, a representation of the group is obtained by exponentiating the representation of the algebra, $D(g(\alpha)) = e^{i\alpha\cdot d(L)}$. More precisely, this provides a representation of that Lie group with the given Lie algebra which

[52] To fix your ideas, think of a d-dimensional hypersurface living in $\mathbb{R}^{d+1}$, e.g., a (hyper)sphere $S^d \subset \mathbb{R}^{d+1}$ defined by the condition $\sum_{a=1}^{d+1} x_a^2 = 1$ on the coordinates x_a of its points.

is also simply connected (such a group always exists). For irreducible representations, this in turn provides representations for the other groups sharing the same algebra, which is generally a projective representation.

To see this more directly, notice that any element of a connected Lie group can be obtained by composing many "small" elements close to the identity, $g = \prod_{n=1}^{N} g_n$, and making these closer and closer to the identity one obtains the composition of infinitely many infinitesimal elements. For example, any finite rotation can be obtained as a sequence of infinitesimal rotations. Then for the representative matrices we have

$$D(g) = \prod_{n=1}^{N} D(g_n) = \prod_{n=1}^{N} D(e^{i\epsilon\alpha_a^{(n)} L_a}) \,, \tag{4.73}$$

with small ϵ and arbitrary $\alpha_a^{(n)}$. Let us now tentatively define

$$D(e^{i\epsilon\alpha_a L_a}) \equiv e^{i\epsilon\alpha_a d(L_a)} \,, \tag{4.74}$$

with a yet to be determined linear mapping $d(L_a)$, i.e., $\alpha_a d(L_a) = d(\alpha_a L_a)$, obeying Eq. (4.71). The $d(L_a)$ will be the representatives of the group generators if D is indeed a representation of the group. For the product of two representative matrices $D(e^{i\epsilon\alpha_a^{(1,2)} L_a})$ we must have (here $A^{1,2} = \alpha_a^{(1,2)} L_a$)

$$\begin{aligned}
&D(e^{i\epsilon A^{(1)}}) D(e^{i\epsilon A^{(2)}}) = D(e^{i\epsilon A^{(1)}} e^{i\epsilon A^{(2)}}) \\
&= D\left(\mathbf{1} + i\epsilon(A^{(1)} + A^{(2)}) - \tfrac{1}{2}\epsilon^2 (A^{(1)\,2} + A^{(2)\,2}) - \epsilon^2 A^{(1)} A^{(2)}\right) \\
&= D\left(\mathbf{1} + i\epsilon(A^{(1)} + A^{(2)}) - \tfrac{1}{2}\epsilon^2 (A^{(1)} + A^{(2)})^2 - \tfrac{1}{2}\epsilon^2 [A^{(1)}, A^{(2)}]\right) \\
&= e^{i\epsilon d\left(A^{(1)} + A^{(2)} + \frac{i}{2}\epsilon [A^{(1)}, A^{(2)}]\right)} = e^{i\epsilon\left(d(A^{(1)}) + d(A^{(2)}) + \frac{i}{2}\epsilon d([A^{(1)}, A^{(2)}])\right)} \\
&= e^{i\epsilon\left(d(A^{(1)}) + d(A^{(2)}) + \frac{i}{2}\epsilon [d(A^{(1)}), d(A^{(2)})])\right)}
\end{aligned} \tag{4.75}$$

where in the last three passages we used the definition Eq. (4.74), linearity of d, and the second part of Eq. (4.71). On the other hand

$$\begin{aligned}
&D(e^{i\epsilon A^{(1)}}) D(e^{i\epsilon A^{(2)}}) = e^{i\epsilon d(A^{(1)})} e^{i\epsilon d(A^{(2)})} \\
&= \left(\mathbf{1} + i\epsilon d(A^{(1)}) - \tfrac{1}{2}\epsilon^2 d(A^{(1)})^2\right) \left(\mathbf{1} + i\epsilon d(A^{(2)}) - \tfrac{1}{2}\epsilon^2 d(A^{(2)})^2\right) \\
&= \left(\mathbf{1} + i\epsilon \left(d(A^{(1)}) + d(A^{(2)})\right) - \tfrac{1}{2}\epsilon^2 \left(d(A^{(1)})^2 + d(A^{(2)})^2\right) - \epsilon^2 d(A^{(1)}) d(A^{(2)})\right) \\
&= \left(\mathbf{1} + i\epsilon \left(d(A^{(1)}) + d(A^{(2)})\right) - \tfrac{1}{2}\epsilon^2 \left(d(A^{(1)}) + d(A^{(2)})\right)^2 - \tfrac{1}{2}\epsilon^2 [d(A^{(1)}), d(A^{(2)})]\right) \\
&= e^{i\epsilon\left(d(A^{(1)}) + d(A^{(2)}) + \frac{i}{2}\epsilon [d(A^{(1)}), d(A^{(2)})]\right)} \,,
\end{aligned} \tag{4.76}$$

so indeed D is a representation. In order to make sure that it is a regular, non-projective representation, one should show that the same $D(g)$ is obtained for a finite transformation g independently of how one builds it out of infinitesimal transformations. This is the case for simply connected groups, while for the non-simply connected ones one can get results differing by a phase, and so $D(g)$ obtained in this way is a projective representation.

It is clear that an irreducible unitary representation of the group yields an irreducible representation of the algebra in terms of Hermitean matrices, and vice versa. Using the notation of Eq. (4.58), we can write the entries of a unitary matrix $D(g)_{n'n}$ as the matrix elements of a unitary operator $U(g)$, i.e., $D(g)_{n'n} = \langle n'|U(g)|n\rangle$ with $\{|n\rangle\}$ a basis of orthonormal vectors, and for infinitesimal g we find

$$D(g)_{n'n} = \delta_{n'n} + i\alpha_a d(L_a)_{n'n} + \ldots = \langle n'|U(g)|n\rangle = \delta_{n'n} + i\alpha_a \langle n'|H(L_a)|n\rangle + \ldots \tag{4.77}$$

with $H(L_a)$ Hermitean operators with matrix elements $d(L_a)_{n'n}$ between basis vectors, so obeying the same commutation relations as the group generators. If all the $D(g)$ leave a subspace invariant, so will the $H(L_a) = -i\partial D(g)/\partial\alpha_a|_{\alpha=0}$; conversely, if the $H(L_a)$ leave a subspace invariant then so do the small transformations $D(g_0) = e^{i\alpha\cdot H(L_a)}$, and so all the transformations.

4.2.3 SO(3) and SU(2)

We now return to our problem, namely finding the unitary (possibly projective) representations of SO(3). Since these are always completely reducible, we just need to find the irreducible (possibly projective) (possibly projective) unitary representations of SO(3).[53] As a manifold, SO(3) is connected but not simply connected, but shares the same Lie algebra with the simply connected group SU(2), i.e., the group of unitary 2×2 complex matrices. The problem of classifying the physically relevant representations of SO(3) then boils down to classifying the irreducible unitary representations of SU(2), and so the irreducible Hermitean representations of the corresponding algebra,[54] which give us projective representations of SO(3). A result of Bargmann [8] tells us that all these are all the projective representations of SO(3).

The fact that SO(3) and SU(2) have the same Lie algebra can be shown explicitly as follows. For SO(3), one looks at the infinitesimal rotations around one of the three coordinate axis. These are of the form $R = \mathbf{1} + i\vec{\theta}\cdot\vec{L}$, with $\vec{\theta}\cdot\vec{L} = \sum_{a=1}^{3}\theta_a \cdot L_a$ and

$$L_1 = -i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad L_2 = -i\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad L_3 = -i\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.78}$$

An explicit calculations shows that

$$[L_a, L_b] = i\varepsilon_{abc}L_c\,, \tag{4.79}$$

with ε_{abc} the totally antisymmetric Levi-Civita symbol. On the other hand, the most general SU(2) matrix is of the form

$$U = u_0\mathbf{1} + i\vec{u}\cdot\vec{\sigma}\,, \qquad u_0^2 + \vec{u}^{\,2} = 1\,, \quad u_{0,1,2,3} \in \mathbb{R}\,, \tag{4.80}$$

where $\sigma_{1,2,3}$ are the Pauli matrices,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.81}$$

[53] One can actually show that for SO(3) all its finite-dimensional representations are equivalent to unitary representations (i.e., they are given by unitary matrices if we choose properly the basis of the representation space $\mathbb{C}^n$).This is a consequence of SO(3) being a *compact group*, i.e., of being compact as a manifold: finite-dimensional representations are always equivalent to a unitary representation.

[54] SU(2) is also a compact group, so classifying its irreducible unitary representations completely classifies all the representations.

Incidentally, this shows that SU(2) is exactly the same manifold as the four-dimensional sphere S^3. Since the Pauli matrices obey $\{\sigma_a, \sigma_b\} = 2\delta_{ab}$, by setting $\vec{u} = \sin\frac{|\alpha|}{2}\hat{\alpha}$ with $\hat{\alpha}$ a unit vector, and $\vec{\alpha} = |\alpha|\hat{\alpha}$, one finds

$$U(\vec{\alpha}) = \cos\frac{|\alpha|}{2}\mathbf{1} + i\sin\frac{|\alpha|}{2}\hat{\alpha}\cdot\vec{\sigma} = e^{i\vec{\alpha}\cdot\frac{\vec{\sigma}}{2}}\,. \tag{4.82}$$

The generators of SU(2) are then $s_a \equiv \frac{\sigma_a}{2}$, that obey

$$[s_a, s_b] = \tfrac{1}{4}[\sigma_a, \sigma_b] = \tfrac{1}{4}2i\varepsilon_{abc}\sigma_c = i\varepsilon_{abc}s_c\,. \tag{4.83}$$

4.2.4 Irreducible representations of SU(2)

Our task is now to find the irreducible Hermitean representations of the algebra Eq. (4.83). This means that we look for a set of Hermitean matrices $d(s_a)$ that obey

$$[d(s_a), d(s_b)] = i\varepsilon_{abc}d(s_c). \tag{4.84}$$

The entries of these matrices are the matrix elements of Hermitean operators $S_a = H(s_a)$ [see Eq. (4.77)], also obeying

$$[S_a, S_b] = i\varepsilon_{abc}S_c, \tag{4.85}$$

between suitable basis vectors $|n\rangle$ to be chosen below, $d(s_a)_{n'n} = \langle n'|S_a|n\rangle$. Notice the distinction: s_a are the group generators; $d(s_a)$ are Hermitean matrices providing a representation of the group algebra; $S_a = H(s_a)$ are Hermitean operators acting on a linear space having the entries of $d(s_a)$ as their matrix elements between basis vectors, which we will then also call representatives of the generators.

Our task is made easier by introducing the *ladder operators* $s_\pm = s_1 \pm is_2$, which obey the commutation relations

$$[s_3, s_\pm] = \pm s_\pm\,, \qquad [s_+, s_-] = 2s_3\,, \tag{4.86}$$

as one can verify explicitly.[55] Instead of looking for the representative operators $S_{1,2,3}$, we look instead for representatives $S_\pm, S_3$ obeying

$$[S_3, S_\pm] = \pm S_\pm\,, \qquad [S_+, S_-] = 2S_3\,, \tag{4.87}$$

from which the desired $S_{1,2}$ are obtained via $S_1 = (S_+ + S_-)/2$ and $S_2 = (S_+ - S_-)/(2i)$. Notice that $S_\pm$ obey the operator relation $S_\pm^\dagger = S_\mp$ [and similarly $d(s_\pm)$ obey the matrix relation $d(s_\pm)^\dagger = d(s_\mp)$].

Another useful object is

$$\vec{s}^{\,2} = s_1^2 + s_2^2 + s_3^2\,, \tag{4.88}$$

[55] A technical but important comment. The objects $s_\pm$ *do not* belong to the algebra of SU(2), since this only involves linear combinations of generators with *real* coefficients. These objects belong instead to the *complexification* of the algebra, where one allows linear combinations of generators with complex coefficients. This can be thought of as a real algebra with twice as many generators, i.e., $\{L_a, iL_a\}$. If one obtains a representation of the complexified algebra, then a representation of the (real) algebra is obtained by restricting it to the subspace corresponding to $\{L_a\}$ only.

which commutes with all the generators:[56]

$$[\vec{s}^{\,2}, s_a] = \sum_b \{s_b, [s_b, s_a]\} = \sum_{b,c} i\varepsilon_{bac}\{s_b, s_c\} = 0\,, \tag{4.89}$$

since the Levi-Civita symbol is antisymmetric and the anticommutator is symmetric in b, c. Such an object commuting with all the generators of a group is called a *Casimir operator*. Correspondingly, we have

$$\vec{S}^{\,2} = S_1^2 + S_2^2 + S_3^2\,, \tag{4.90}$$

built out of the representatives that we are looking for, which is the representative of the Casimir operator. Since by construction the S_a obey the same commutation relations as the s_a, one has that $[\vec{S}^{\,2}, S_a] = 0$. Using the commutation relations one finds

$$\vec{s}^{\,2} = s_- s_+ + s_3^2 + s_3 = s_+ s_- + s_3^2 - s_3\,, \tag{4.91}$$

and similarly for $\vec{S}^{\,2}$,

$$\vec{S}^{\,2} = S_- S_+ + S_3^2 + S_3 = S_+ S_- + S_3^2 - S_3\,. \tag{4.92}$$

We now proceed with the construction. We are interested in finite-dimensional Hermitean representations, i.e., in finding suitable Hermitean matrices acting on some finite-dimensional complex space. Since S_a are Hermitean we can surely diagonalise one of them, say, S_3 [we cannot simultaneously diagonalise any other of the S_a due to the commutation relations Eq. (4.84)], and use the corresponding eigenvectors as the basis of our representation space. Given an eigenvector $|m\rangle$ of S_3 of eigenvalue m, we have from Eq. (4.87)

$$S_3 S_\pm |m\rangle = ([S_3, S_\pm] + S_\pm S_3)|m\rangle = (m \pm 1) S_\pm |m\rangle\,, \tag{4.93}$$

i.e., $S_\pm|m\rangle$ is another eigenvector of S_3 (or zero). Similarly, any of the $S_\pm^k|m\rangle$ is an eigenvector of S_3 (or zero). Since the dimension of our representation space is finite, there must be some value of k for which $S_+^k|m\rangle = 0$ (we would have an infinite sequence of eigenvectors otherwise), and so there is a vector $|s,s\rangle \equiv S_+^{k-1}|m\rangle$ obeying $S_3|s,s\rangle = s|s,s\rangle$ for some s, and moreover $S_+|s,s\rangle = 0$. From Eq. (4.92) we also find $\vec{S}^2|s,s\rangle = s(s+1)|s,s\rangle$. If we now apply S_- to $|s,s\rangle$ we get an eigenvector of S_3 of eigenvalues $s-1$, and of $\vec{S}^2$ with the same eigenvalue $s(s+1)$, since any linear combinations of generators and so in particular the ladder operators commute with it. Iterating the procedure, we get eigenvectors of ever decreasing eigenvalue. This procedure must again stop at some point. To see where, notice that if we normalise $|s,s\rangle$ to 1, $\langle s,s|s,s\rangle = 1$, then in general $S_-^k|s,s\rangle \propto |s, s-k\rangle$, where $|s,m\rangle$ are normalised to 1 – but with a proportionality factor diffent from 1. The scalar product here is the one naturally associated with $\mathbb{C}^n$. Now,

$$\langle s,m|S_\pm S_\mp|s,m\rangle = \langle s,m|\vec{S}^{\,2} - S_3^2 \pm S_3|s,m\rangle = s(s+1) - m(m\mp 1)\,, \tag{4.94}$$

and so the descent will stop when $s(s+1) - m(m-1) = 0$, since we are not generating a new eigenvector. This is solved by $m = s+1$, which cannot be;[57] and by $m = -s$, meaning that

[56] Also $\vec{s}^{\,2}$ does not belong to the Lie algebra, but rather to the universal enveloping algebra – whatever that thing is. Similarly, the operator $\vec{S}^{\,2}$ is not the mapping of $\vec{s}^{\,2}$ via H, $\vec{S}^{\,2} = [H(\vec{s})]^{\,2} \neq H(\vec{s}^{\,2})$, since H is defined for real linear combinations $X^a s_a$ by linear extension from $H(s_a)$.

[57] A vector $|m\rangle$ with $S_3|m\rangle = m|\psi\rangle$ necessarily obeys $\langle m|\vec{S}^{\,2}|m\rangle \geq m(m+1)$, so if $m = s+1$ it cannot be also an eigenvector of $\vec{S}^{\,2}$ with eigenvalue $s(s+1)$.

$S_-|s,-s\rangle = 0$. There is then a total of $2s+1$ basis eigenvectors, with corresponding eigenvalues of S_3 equal to $s, s-1, \ldots, -s+1, -s$, and since this number must be a non-negative integer then s can be only a non-negative integer or half-integer. Equation (4.94) can be used also to iteratively normalise the vectors obtained applying the ladder operator S_-,

$$|s,m-1\rangle \equiv \frac{1}{\sqrt{s(s+1)-m(m-1)}} S_-|s,m\rangle\,. \tag{4.95}$$

Here we have made a definite choice of phase, corresponding to the request that S_- has positive-definite matrix elements between basis vectors. This is known as the *Condon-Shortley convention.*

The representation we have just obtained is irreducible. In fact, if it were not then there would be some invariant subspace, and since by repeatedly applying S_+ to any of its vectors we would obtain a new vector of the invariant subspace, then we must find some $|\psi'\rangle$ such that $S_+|\psi'\rangle = \alpha|\psi'\rangle$, i.e., an eigenvector of S_+ (possibly with eigenvalue 0), so that the sequence stops. But then

$$\alpha\langle s,-s|\psi'\rangle = \langle s,-s|S_+|\psi'\rangle = 0\,, \tag{4.96}$$

so either $\alpha = 0$ or $\langle s,-s|\psi'\rangle = 0$. In the first case we find for any $s < m$ that $0 = \langle s,s|S_+^{s-m}|\psi'\rangle \propto \langle s,m|\psi'\rangle$ so that $|\psi'\rangle$ must be orthogonal to all $|s,m\rangle$ except possibly $|s,s\rangle$, and so can only be $|s,s\rangle$ itself – in which case the invariant subspace is the whole representation space; or the zero vector. In the second case,

$$\begin{aligned} 0 &= \langle s,-s|\vec{S}^2|\psi'\rangle = \langle s,-s|S_-S_+ + S_3^2 + S_3|\psi'\rangle = \langle s,-s|S_-S_+|\psi'\rangle \\ &= \alpha\langle s,-s|S_-|\psi'\rangle \propto \langle s,-s+1|\psi'\rangle\,. \end{aligned} \tag{4.97}$$

Repeating the argument for $\langle s,-s+1|$, then $\langle s,-s+2|$, and so on, we find that $|\psi'\rangle$ is orthogonal to all basis vectors, and so it must be the zero vector.

Summarising, we have found that the irreducible representations of SU(2) are labelled by an integer or half-integer number s, that we will call the *spin* of the representation. The dimension of the representation is $2s+1$, and the representatives $S_{\pm,3}$ of $s_{\pm,3}$ act on the basis elements $|s,m\rangle$ to give

$$\begin{aligned} S_3|s,m\rangle &= m|s,m\rangle\,, \\ S_\mp|s,m\rangle &= \sqrt{s(s+1)-m(m\mp1)}|s,m\mp1\rangle\,. \end{aligned} \tag{4.98}$$

Finally, the *total spin* operator $\vec{s}^2$ is represented simply by a constant multiple of the identity, $\vec{s}^2 \to \vec{S}^2 = s(s+1)\mathbf{1}$,

$$\vec{S}^2|s,m\rangle = s(s+1)|s,m\rangle\,. \tag{4.99}$$

The matrices providing the sought-after representations of SU(2) are then

$$\begin{aligned} d^{(s)}(s_3)_{m'm} &= \langle s,m'|S_3|s,m\rangle = m\delta_{m'm}\,, \\ d^{(s)}(s_\mp)_{m'm} &= \langle s,m'|S_3|s,m\rangle = \delta_{m',m\mp1}\sqrt{s(s+1)-m(m\mp1)}\,, \\ ([d^{(s)}(\vec{s})]^2)_{m'm} &= s(s+1)\delta_{m'm}\,. \end{aligned} \tag{4.100}$$

The value of s, or equivalently the eigenvalue of $\vec{S}^2$, entirely characterises the representation. The representations of lowest dimension are listed in Table 8. Let us see what a few of these representations look like explicitly (see Table 9).

s	$2s+1$	$s(s+1)$
0	1	0
$\frac{1}{2}$	2	$\frac{3}{4}$
1	3	2
$\frac{3}{2}$	4	$\frac{15}{4}$

Table 8: Irreducible representations of SU(2) of lowest dimension.

The simplest representation is the 1-dimensional, $s = 0$ representation. This consists in mapping every $U \in \mathrm{SU}(2)$ to $D_0(U) = 1$, and is generated by $d_0(\vec{s}) = 0$. This is the *trivial* representation; its single state is called a singlet.

The second simplest representation of SU(2) is the 2-dimensional representation provided by the group elements themselves, since they are matrices, $D_{\frac{1}{2}}(U) = U$, and has as representatives of the generators the generators themselves, $d_{\frac{1}{2}}(\vec{s}) = \vec{s} = \frac{\vec{\sigma}}{2}$. This is the *defining* of *fundamental* representation. The vectors

$$\begin{pmatrix}1\\0\end{pmatrix} = |\tfrac{1}{2}, +\tfrac{1}{2}\rangle\,, \qquad \begin{pmatrix}0\\1\end{pmatrix} = |\tfrac{1}{2}, -\tfrac{1}{2}\rangle\,, \tag{4.101}$$

are the eigenvectors of s_3, with

$$s_\mp|\tfrac{1}{2}, \pm\tfrac{1}{2}\rangle = \tfrac{1}{2}(\sigma_1 \pm i\sigma_2)|\tfrac{1}{2}, \pm\tfrac{1}{2}\rangle = |\tfrac{1}{2}, \mp\tfrac{1}{2}\rangle\,. \tag{4.102}$$

The eigenvectors form a *doublet*. Another 2-dimensional representation is the *complex-conjugate representation*, $D_{\bar{\frac{1}{2}}}(U) = U^*$, generated by $d_{\bar{\frac{1}{2}}}(\vec{s}) = -\frac{\vec{\sigma}^*}{2}$.[58] Since $\vec{\sigma}^* = -\sigma_2\vec{\sigma}^*\sigma_2$, in the specific case of SU(2) the complex-conjugate representation is equivalent to the fundamental; this is in general not the case for other groups. Notice that since $d_{\bar{\frac{1}{2}}}(s_3) = -s_3$, the vectors in Eq. (4.101) are still eigenvectors in the complex conjugate representation but with opposite eigenvalues,

$$\begin{aligned} d_{\bar{\frac{1}{2}}}(s_3)\begin{pmatrix}1\\0\end{pmatrix} &= -\frac{\sigma_3}{2}\begin{pmatrix}1\\0\end{pmatrix} = -\frac{1}{2}\begin{pmatrix}1\\0\end{pmatrix}\,,\\ d_{\bar{\frac{1}{2}}}(s_3)\begin{pmatrix}0\\1\end{pmatrix} &= -\frac{\sigma_3}{2}\begin{pmatrix}0\\1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}0\\1\end{pmatrix}\,. \end{aligned} \tag{4.103}$$

On the other hand, $d_{\bar{\frac{1}{2}}}(s_\pm) = -\frac{1}{2}(\sigma_1^* \pm i\sigma_2^*) = -\frac{1}{2}(\sigma_1 \mp i\sigma_2)$, and so

$$d_{\bar{\frac{1}{2}}}(s_-)\begin{pmatrix}0\\1\end{pmatrix} = -\tfrac{1}{2}(\sigma_1 + i\sigma_2)\begin{pmatrix}0\\1\end{pmatrix} = -\begin{pmatrix}1\\0\end{pmatrix}\,. \tag{4.104}$$

To satisfy the Condon-Shortley convention (S_- has non-negative matrix elements), we need to include a minus sign in the defintion of one of the two basis vectors, e.g.,

$$\begin{pmatrix}0\\1\end{pmatrix} = -|\bar{\tfrac{1}{2}}, +\tfrac{1}{2}\rangle\,, \qquad \begin{pmatrix}1\\0\end{pmatrix} = |\bar{\tfrac{1}{2}}, -\tfrac{1}{2}\rangle\,, \tag{4.105}$$

where $\bar{\frac{1}{2}}$ denotes the fact that this is the complex-conjugate representation.

[58] This is a representation since $D_{\bar{\frac{1}{2}}}\big(U(\vec{\alpha})\big)D_{\bar{\frac{1}{2}}}\big(\big(U\vec{\beta}\big)\big) = U(\vec{\alpha})^*U(\vec{\beta})^* = \big(U(\vec{\alpha})U(\vec{\beta})\big)^* = D_{\bar{\frac{1}{2}}}\big(\big(U(\vec{\alpha})U(\vec{\beta})\big)\big)$.

A Lie group can always be represented on its own Lie algebra. For matrix groups, this is achieved by the following linear mappings of the vector space spanned by the generators $\vec{s} = \frac{\vec{\sigma}}{2}$ to itself,

$$D_1(U) = \mathrm{Ad}_U X \equiv UXU^\dagger\,, \qquad X = X^a s_a\,, \quad X^a \in \mathbb{R}\,. \tag{4.106}$$

This is the *adjoint representation*. Since the algebra of SU(2) is a 3-dimensional space, the adjoint is a 3-dimensional representation; in general it has the same dimension as the algebra of the Lie group. For infinitesimal transformations $U = \mathbf{1} + i\alpha_a s_a$,

$$D_1(U)X = \mathrm{Ad}_U X = X + i\alpha_a[s_a, X] = X + i\alpha_a \mathrm{ad}_{s_a} X = X + i\alpha_a d_1(s_a)\,. \tag{4.107}$$

For any fixed element A of the algebra, ad_A is a linear operator acting on the algebra itself,

$$\mathrm{ad}_A X \equiv [A, X]\,. \tag{4.108}$$

One has in components

$$\mathrm{ad}_{s_a} X = (\mathrm{ad}_{s_a} X)^c s_c = s_c (\mathrm{ad}_{s_a})_{cb} X^b = X^b [s_a, s_b] = X^b i\varepsilon_{abc} s_c = s_c(-i\varepsilon_{acb})X^b\,, \tag{4.109}$$

i.e., $(\mathrm{ad}_{s_a})_{cb} = -i\varepsilon_{acb} \equiv (T_a^{(A)})_{cb}$. Using the Jacobi identity,

$$[A,[B,C]] + [C,[A,B]] + [B,[C,A] = 0\,, \tag{4.110}$$

valid for any three $n \times n$ matrices A, B, C, one shows that for every X,

$$\begin{aligned}[\mathrm{ad}_{s_a}, \mathrm{ad}_{s_b}]X &= [s_a,[s_b,X]] - [s_b,[s_a,X]] = [[s_a,s_b],X] + [s_b,[s_a,X]] - [s_b,[s_a,X]] \\ &= i\varepsilon_{abc}[s_c, X] = i\varepsilon_{abc}\mathrm{ad}_{s_c} X\,,\end{aligned} \tag{4.111}$$

and so

$$[\mathrm{ad}_{s_a}, \mathrm{ad}_{s_b}] = i\varepsilon_{abc}\mathrm{ad}_{s_c}\,, \tag{4.112}$$

or in terms of its matrix components

$$[T_a^{(A)}, T_b^{(A)}] = i\varepsilon_{abc} T_c^{(A)}\,. \tag{4.113}$$

From the definition of the $T_a^{(A)}$

$$T_1^{(A)} = -i\begin{pmatrix}0&0&0\\0&0&1\\0&-1&0\end{pmatrix}\,, \quad T_2^{(A)} = -i\begin{pmatrix}0&0&-1\\0&0&0\\1&0&0\end{pmatrix}\,, \quad T_3^{(A)} = -i\begin{pmatrix}0&1&0\\-1&0&0\\0&0&0\end{pmatrix}\,, \tag{4.114}$$

which are precisely the generators of SO(3), Eq. (4.78).[59] The eigenvectors of s_3 in this representation, i.e., the eigenvectors of $T_3^{(A)}$, are read off from the commutation relations Eq. (4.86): since $[s_3, s_\pm] = \pm s_\pm$,

$$T_3^{(A)} s_\pm = \mathrm{ad}_{s_3} s_\pm = [s_3, s_\pm] = \pm s_\pm\,, \qquad T_3^{(A)} s_3 = \mathrm{ad}_{s_3} s_3 = [s_3, s_3] = 0\,, \tag{4.115}$$

so the eigenvectors are the algebra elements[60] s_+, s_3, s_-, with eigenvalues $1, 0, -1$, respectively. These form a *triplet*.

[59] These differ from the ones obtained with the construction described above, but are unitarily equivalent to them, i.e., they are related by a unitary change of basis.

[60] The elements $s_\pm$ are actually part of the complexification of the group algebra, where linear combinations of the generators with complex coefficients are allowed. Since here we are treating the algebra as a vector space where the representation acts as a set of complex matrices, there is no problem in doing this.

s	dimension	generators rep.	group rep.	representation
0	1	$d_0(\vec{s}) = 0$	$D_0(U) = 1$	trivial
$\frac{1}{2}$	2	$d_{\frac{1}{2}}(\vec{s}) = \frac{\vec{\sigma}}{2}$	$D_{\frac{1}{2}}(U) = U$	fundamental/defining
$\frac{1}{2}$	2	$d_{\bar{\frac{1}{2}}}(\vec{s}) = -\frac{\vec{\sigma}^*}{2}$	$D_{\bar{\frac{1}{2}}}(U) = U^*$	complex conjugate (equiv. to fundamental)
1	3	$d_1(\vec{s}) = \mathrm{ad}_{\vec{s}}$	$D_1(U) = \mathrm{Ad}_U$	adjoint

Table 9: Explicit form of the lowest-dimensional irreducible representations of SU(2).

To conclude this subsection, we go back to our original problem, which was to understand the finite-dimensional Hilbert space of states of a massive particle at rest. We have shown that this space supports a unitary representation of SU(2). This in general could be reducible, but there would be no reason for all its irreducible components to carry exactly the same rotation-invariant quantum numbers, like, e.g., the mass. We then expect one value of the mass to correspond to a single irreducible representation. We are then led to characterise particles in terms of two quantities: their mass m, and their spin s, characterising the irreducible representation of the rotation group under which the states in the rest frame transform – i.e., the behaviour of the particle under rotations.

4.2.5 Composition of representations

If we consider the states of two electron, the corresponding wave functions $\psi_1\psi_2$ have now four components, coming from the direct product of the wave functions associated with each electron. The representation under which such states transform results from the composition of the individual $s = \frac{1}{2}$ representations:

$$(\psi_1)_i(\psi_2)_j \to [D(U)\psi_1]_i[D(U)\psi_2]_j = D(U)_{ii'}D(U)_{jj'}(\psi_1)_{i'}(\psi_2)_{j'} \,. \tag{4.116}$$

The matrices $\mathcal{D}_{ij;i'j'}(U) = D(U)_{ii'}D(U)_{jj'}$ combining the components $(\Psi_{12})_{i'j'} = (\psi_1)_{i'}(\psi_2)_{j'}$ of the total wave function clearly still provide a representation of SU(2),

$$\begin{aligned}\mathcal{D}_{ij;i'j'}(U_2U_1) &= D(U_2U_1)_{ii'}D(U_2U_1)_{jj'} = D(U_2)_{ii''}D(U_1)_{i''i'}D(U_2)_{jj''}D(U_1)_{j''j'} \\ &= D(U_2)_{ii''}D(U_2)_{jj''}D(U_1)_{i''i'}D(U_1)_{j''j'} = \mathcal{D}_{ij;i''j''}(U_2)\mathcal{D}_{i''j'';i'j'}(U_1)\,,\end{aligned} \tag{4.117}$$

but not necessarily an irreducible one. However, since $\mathcal{D}$ is a unitary representation because $D(U)$ are unitary, one can decompose it into the direct sum of irreducible representations.[61]

Let us denote the basis vectors of the $2 \times 2 = 4$ dimensional space of wave functions as

$$|\tfrac{1}{2}, m_1\rangle \otimes |\tfrac{1}{2}, m_2\rangle\,, \qquad m_{1,2} = \pm\tfrac{1}{2}\,. \tag{4.118}$$

Denote $\vec{S}^{(1)} = \vec{S} \otimes \mathbf{1}$ and $\vec{S}^{(2)} = 1 \otimes \vec{S}$, with $\vec{S}$ the representatives of the group generators in the $s = \frac{1}{2}$ representation, corresponding to $D(U)$ in Eq. (4.116). The action of $\vec{S}^{(1,2)}$ on the factor $|\frac{1}{2}, m_{2,1}\rangle$ of the basis vectors is trivial. The combination $\vec{S}^{\,\mathrm{tot}} = \vec{S}^{(1)} + \vec{S}^{(2)}$ provides another

[61] More generally, since any finite-dimensional representation of SU(2) is equivalent to a unitary representation which is completely reducible, any finite-dimensional representation of SU(2) is always completely reducible.

representation on the space spanned by $|\frac{1}{2}, m_{1,2}\rangle$, corresponding precisely to the representatives of the group generators obtained from $\mathcal{D}(U)$. Clearly, $S_3^{\rm tot}$ is diagonal in the basis Eq. (4.118), with eigenvalues equal to the sum $m = m_1 + m_2$ of the eigenvalues. The maximal value of m is $m = \frac{1}{2} + \frac{1}{2} = 1$, obtained from the vector annihilated by both $S_+^{(1)}$ and $S_+^{(2)}$, and so by $S_+^{\rm tot}$. This vector must then be an eigenvector of $(\vec{S}^{\rm tot})^2$, and the corresponding eigenvalue $s(s+1)$ is found to be 2, i.e., the eigenvalue of

$$|\vec{S}^{\rm tot}| \equiv \frac{-1 + \sqrt{1 + (\vec{S}^{\rm tot})^2}}{2} \tag{4.119}$$

is $s = 1$. If we now apply the construction of the previous subsection, we will find an irreducible representation with $s = 1$:

$$\begin{array}{lllll} & & |\vec{S}^{\rm tot}|: & S_3^{\rm tot}: \\ & |\frac{1}{2},\frac{1}{2}\rangle \otimes |\frac{1}{2},\frac{1}{2}\rangle & s=1 & m=1\,, \\ \text{apply } S_-^{\rm tot} & \downarrow & & \\ & |\frac{1}{2},-\frac{1}{2}\rangle \otimes |\frac{1}{2},\frac{1}{2}\rangle + |\frac{1}{2},\frac{1}{2}\rangle \otimes |\frac{1}{2},-\frac{1}{2}\rangle & s=1 & m=0\,, \\ \text{apply } S_-^{\rm tot} & \downarrow & & \\ & |-\frac{1}{2},-\frac{1}{2}\rangle \otimes |\frac{1}{2},\frac{1}{2}\rangle + |-\frac{1}{2},\frac{1}{2}\rangle \otimes |\frac{1}{2},-\frac{1}{2}\rangle & s=1 & m=-1\,. \end{array} \tag{4.120}$$

Normalising each vector to 1 we find

$$\begin{aligned} |\tfrac{1}{2},\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle &= |1,1\rangle\,, \\ \tfrac{1}{\sqrt{2}}\left(|\tfrac{1}{2},-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle + |\tfrac{1}{2},\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle\right) &= |1,0\rangle\,, \\ |\tfrac{1}{2},-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle &= |1,-1\rangle\,. \end{aligned} \tag{4.121}$$

The three-dimensional subspace spanned by these three vectors is left invariant by the $\vec{S}^{\rm tot}$, but contains no further invariant subspace. There is one more independent vector: if we look at Eq. (4.120), we readily find that there is another combination of $|\frac{1}{2},-\frac{1}{2}\rangle \otimes |\frac{1}{2},\frac{1}{2}\rangle$ and $|\frac{1}{2},\frac{1}{2}\rangle \otimes |\frac{1}{2},-\frac{1}{2}\rangle$ orthogonal to $|1,0\rangle$, namely

$$\tfrac{1}{\sqrt{2}}\left(|\tfrac{1}{2},-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle - |\tfrac{1}{2},\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle\right) = |0,0\rangle\,, \tag{4.122}$$

which gives an irreducible representation with $s = 0$. We have then showed that the direct product of two $s = \frac{1}{2}$ representations equals the direct sum of the $s = 1$ and the $s = 0$ representations, i.e., $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$. This can be expressed in term of the dimensions of the representations as $\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3}$.

The same game can be played composing more general representations. For example, composing $s = \frac{1}{2}$ and $s = 1$ we find, starting from the largest eigenvalue of $S_3^{\rm tot}$, which is $m = \frac{1}{2} + 1 = \frac{3}{2}$, and applying $S_-^{\rm tot}$ repeatedly (normalising the vectors along the way)

$$\begin{aligned} |\tfrac{1}{2},\tfrac{1}{2}\rangle \otimes |1,1\rangle &= |\tfrac{3}{2},\tfrac{3}{2}\rangle\,, \\ \tfrac{1}{\sqrt{3}}\left(|\tfrac{1}{2},-\tfrac{1}{2}\rangle \otimes |1,1\rangle + \sqrt{2}|\tfrac{1}{2},\tfrac{1}{2}\rangle \otimes |1,0\rangle\right) &= |\tfrac{3}{2},\tfrac{1}{2}\rangle\,, \\ \tfrac{1}{\sqrt{3}}\left(\sqrt{2}|\tfrac{1}{2},-\tfrac{1}{2}\rangle \otimes |1,0\rangle + |\tfrac{1}{2},\tfrac{1}{2}\rangle \otimes |1,-1\rangle\right) &= |\tfrac{3}{2},-\tfrac{1}{2}\rangle\,, \\ |\tfrac{1}{2},-\tfrac{1}{2}\rangle \otimes |1,-1\rangle &= |\tfrac{3}{2},-\tfrac{3}{2}\rangle\,, \end{aligned} \tag{4.123}$$

since $|\frac{1}{2},\frac{1}{2}\rangle\otimes|1,1\rangle$ is annihilated by $S_+^{\rm tot}$ and so must be an eigenvector of $|\vec{S}^{\,\rm tot}|$ with eigenvalue equal to $s=m=\frac{3}{2}$. There must be a second vector orthogonal to $|\frac{3}{2},\frac{1}{2}\rangle$ built out of the same basis vectors and so with eigenvalue of $S_3^{\rm tot}$ equal to $\frac{1}{2}$, and moreover annihilated by $S_+^{\rm tot}$ (there is only one vector with eigenvalue of $S_3^{\rm tot}$ equal to $\frac{3}{2}$). This must then be an eigenvector of $|\vec{S}^{\,\rm tot}|$ with eigenvalue $s=\frac{1}{2}$. Identifying this vector and using again $S_-^{\rm tot}$ we find

$$\begin{aligned}\tfrac{1}{\sqrt{3}}\left(\sqrt{2}|\tfrac{1}{2},-\tfrac{1}{2}\rangle\otimes|1,1\rangle-|\tfrac{1}{2},\tfrac{1}{2}\rangle\otimes|1,0\rangle\right)&=|\tfrac{1}{2},\tfrac{1}{2}\rangle\,,\\ \tfrac{1}{\sqrt{3}}\left(|\tfrac{1}{2},-\tfrac{1}{2}\rangle\otimes|1,0\rangle-\sqrt{2}|\tfrac{1}{2},\tfrac{1}{2}\rangle\otimes|1,-1\rangle\right)&=|\tfrac{1}{2},-\tfrac{1}{2}\rangle\,.\end{aligned}\tag{4.124}$$

There are no more independent vectors, and we have shown that $\frac{1}{2}\otimes 1=\frac{1}{2}\oplus\frac{3}{2}$, or in term of the dimensions $\mathbf{2}\otimes\mathbf{3}=\mathbf{2}\oplus\mathbf{4}$. In general, one finds for the composition of representations s_1 and s_2

$$s_1\otimes s_2=|s_1-s_2|\oplus|s_1-s_2|+1\oplus s_1+s_2-1\oplus s_1+s_2\,,\tag{4.125}$$

each appearing with multiplicity 1.

5 Strong interactions

In this Section we discuss the consequences of the approximate flavour symmetries of QCD for strong interactions. These symmetries were known before the discovery of QCD, to which they led.

5.1 Isospin symmetry of the strong interactions

In 1932 Chadwick discovered the neutron, thus solving the puzzle of the mismatch between the mass and charge of the nuclei. In fact, it turned out that the neutron and the proton have very similar masses: $m_n=939.57$ MeV and $m_p=938.28$ MeV, so that $(m_n-m_p)/m_p\simeq 0.0014$. While the nuclear charge is e times the number of protons in the nucleus, the nuclear mass is very accurately m_p times the number of protons and neutrons.

The smallness of the mass difference between proton and neutron led Heisenberg, in the same year 1932, to propose that these particles are actually two different states of the same particle, the *nucleon*, and that they are affected in the same way by the strong interactions. More precisely, he assumed that strong interactions were *exactly* invariant under the exchange of proton and neutron; the small mass difference he attributed to electromagnetic effects. We know now that in fact this symmetry would be approximate even if electromagnetic interactions were switched off, and that an important role in establishing the mass difference between proton and neutron is played by the mass difference between the up and down quarks: in fact, if it were only for electromagnetism we would have $m_p>m_n$ – with catastrophic consequences.

The symmetry that Heisenberg assumed for strong interactions was in fact more than just that under exchange of proton and neutron. If p and n are two states of the nucleon (N), the possible state vectors of the latter are $|\vec{p},s_z;p\rangle$, $|\vec{p},s_z;n\rangle$, and due to the superposition principle also any linear combination of these two. At fixed momentum and spin, the state space of the nucleon is two-dimensional, and so we can assign to p and n the two-component basis vectors

$$p=\begin{pmatrix}1\\0\end{pmatrix}\,,\qquad n=\begin{pmatrix}0\\1\end{pmatrix}\,,\tag{5.1}$$

and represent the most general (normalised) nucleon state as

$$N(\alpha,\beta) = \alpha p + \beta n = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \qquad \alpha,\beta\in\mathbb{C}\,, \qquad |\alpha|^2+|\beta|^2 = 1\,. \tag{5.2}$$

This two-dimensional space corresponds to an *internal* quantum number, i.e., unrelated to space-time. The assumption is now that any of these states looks the same to strong interactions, or, in mathematical terms, that strong interactions are invariant under a general unitary transformation of the nucleon state. Except for irrelevant overall phase redefinitions, this means that strong interactions are assumed to have a symmetry under SU(2) transformations of the nucleon state. This is the *isospin* symmetry. Like any other symmetry that applies to an internal quantum number, isospin symmetry is called an *internal symmetry*.

Isospin and SU(2) Although not exact,[62] isospin symmetry turns out to be a very good approximate symmetry of strong interactions (we will soon see why). While the origin of this symmetry was unknown at the time, its consequences were clearly observed in experiments. An immediate application was to the study of atomic nuclei with the same total number of protons and neutrons, i.e., isotopic (or more precisely *isobaric*) nuclei, which are related by SU(2) transformations. As more and more hadrons were discovered in the following years, isospin symmetry was extended further by postulating that strong interactions possessed a more general internal SU(2) symmetry, of which the symmetry between protons and neutrons was just a manifestation.

Let us now assume that in an ideal world without electromagnetic and weak interactions, and where protons and neutrons have exactly the same mass, the strong interactions possess an exact SU(2) symmetry. Since any superposition of proton and neutron would look the same to the strong interactions, calling one of them the proton and the orthogonal combination the neutron is a matter of the observer's conventions, and any choice would be equivalent to any other. This means that an SU(2) transformation is just a change of conventions that does not change the physics - i.e., it is a symmetry.

Since SU(2) is a continuous group, the SU(2) symmetry transformations mentioned above are implemented by unitary operators U on the Hilbert space $\mathcal{H}$ of strongly interacting matter. While entirely unrelated to spin in physical origin, isospin displays exactly the same mathematical structure, which we can take wholesale from the previous section. In particular, the operators U provide a representation of SU(2) on $\mathcal{H}$, that will break up into the irreducible representations discussed there. The effects of strong interactions being unchanged by SU(2) transformations means that they leave the strong Hamiltonian H_{strong} invariant, or equivalently that they commute with it. Although the group is the same, we typically use a different notation for spin and isospin. The SU(2) matrices corresponding to isospin transformations are written as $U(\vec{\alpha}) = e^{i\vec{\alpha}\cdot\vec{I}} = e^{i\vec{\alpha}\cdot\frac{\vec{\sigma}}{2}}$, with σ_a the familiar Pauli matrices. In the literature one often finds them denoted as τ_a in the context of isospin transformations. The generators are denoted as I_a, $a=1,2,3$ and obey the familiar commutation relations

$$[I_a, I_b] = i\varepsilon_{abc} I_c\,. \tag{5.3}$$

[62] In fact, if it were exact, we would not be able to distinguish protons and neutrons at all: it is precisely the breaking, although small, of isospin symmetry that allows us to separate them.

The corresponding unitary operators on $\mathcal{H}$ read $U(\vec{\alpha}) = e^{i\vec{\alpha}\cdot\vec{I}}$, where $\vec{I}$ are now Hermitian operators and obey the commutation relations Eq. (5.3).[63] Invariance of H_{strong} under isospin rotations amounts to asking $[U(\vec{\alpha}), H_{\text{strong}}] = 0$ for arbitrary $\vec{\alpha}$, which is equivalent to $[\vec{I}, H_{\text{strong}}] = 0$. In the Heisenberg picture, this corresponds to the operators $\vec{I}(t) = e^{iHt}\vec{I}(0)e^{-iHt}$ being time-independent, $d\vec{I}(t)/dt = i[H_{\text{strong}}, \vec{I}(t)] = 0$, i.e., conservation of isospin in strong interactions. The internal nature of isospin symmetry means that isospin transformations commute also with translations and rotations, and so $\vec{I}$ commute with the energy, momentum and angular momentum operators, so that $\vec{I}^2$ and one of its components, say, I_3, can be diagonalised together with $E^2 - \vec{p}^2 = m^2$, $\vec{p}$, $\vec{J}^2$, and J_3.

There are two important consequences of isospin symmetry, that can be straighforwardly tested in experiments:

- the spectrum of the theory is organised in degenerate isospin multiplets, which form the bases of irreducible representations of SU(2);
- isospin is conserved in dynamical hadronic processes, i.e., in decay and scattering processes, leading to relations among decay widths and cross sections.

Mass multiplets and group representations The most evident consequence of isospin symmetry is the existence of multiplets of hadrons with nearly-degenerate masses. Degenerate multiplets are in fact expected in the presence of an exact symmetry. As we recalled above, the SU(2) representation on the whole of $\mathcal{H}$ breaks up into irreducible representations that affect only subspaces of $\mathcal{H}$ that are left invariant by isospin transformations. The irreducible representations are entirely characterised by the eigenvalue $I(I+1)$ of $\vec{I}^2$, which is constant in each invariant subspace, and determines the dimension $2I+1$ of the representation. Since isospin commutes with energy, momentum and angular momentum, which in the rest frame of a particle is just its spin, one has in particular

$$[\vec{I}, p^2] = 0\,, \qquad [\vec{I}, \vec{s}^2] = 0\,, \tag{5.4}$$

and so all these states have the same mass m and spin s. There are then $2I+1$ set of linearly independent particle states $|m, \vec{p}; s, s_z; I, I_3\rangle$ with the same mass and spin, as well as $\vec{p}$ and s_3, forming degenerate particle multiplets. As soon as isospin is broken, as it actually is in nature, these states are expected to display slightly different masses as well as further quantum numbers, which makes them different particles in all respects. One then expects to observe nearly-degenerate multiplets of hadrons.[64] Let us discuss a few examples (see Table 10).

We started from the proton and the neutron, which form a doublet (p, n) of nearly degenerate particles. In the language of SU(2) representations, they are the basis of a two-dimensional, $I = \frac{1}{2}$ representation. We choose protons and neutrons to be in the fundamental representation, and their corresponding antiparticles $\bar{p}$ and $\bar{n}$ to be in the antifundamental representation; to keep the charge-conjugation transformation simple we assign the same two-dimensional representative

[63] With a little abuse of notation we use the same symbol for the SU(2) matrices and generators, and their representatives on $\mathcal{H}$.

[64] The possible degeneracies are $2I+1$ with I integer or half-integer, so any number is possible in principle.

	I	I_3	Q	B	S	Y	mass
p	$\frac{1}{2}$	$\frac{1}{2}$	1	1	0	1	938.28 MeV
n	$\frac{1}{2}$	$-\frac{1}{2}$	0	1	0	1	939.57 MeV
π^+	1	1	1	0	0	0	139.57 MeV
π^0	1	0	0	0	0	0	134.98 MeV
π^-	1	-1	-1	0	0	0	139.57 MeV
Σ^+	1	1	1	1	-1	0	1.189 GeV
Σ^0	1	0	0	1	-1	0	1.193 GeV
Σ^-	1	-1	-1	1	-1	0	1.19 GeV
K^+	$\frac{1}{2}$	$\frac{1}{2}$	1	0	1	1	493.7 MeV
K^0	$\frac{1}{2}$	$-\frac{1}{2}$	0	0	1	1	497.6 MeV
$\bar{K}^0$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	-1	-1	497.6 MeV
K^-	$\frac{1}{2}$	$-\frac{1}{2}$	-1	0	-1	-1	493.7 MeV
Δ^{++}	$\frac{3}{2}$	$\frac{3}{2}$	2	1	0	1	1.232 GeV
Δ^+	$\frac{3}{2}$	$\frac{1}{2}$	1	1	0	1	1.232 GeV
Δ^0	$\frac{3}{2}$	$-\frac{1}{2}$	0	1	0	1	1.232 GeV
Δ^-	$\frac{3}{2}$	$-\frac{3}{2}$	-1	1	0	1	1.232 GeV

Table 10: Isospin multiplets of hadrons.

in isospin space to particle and antiparticle.[65] We then have for nucleon and antinucleon states

$$\begin{aligned} p &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |N; \tfrac{1}{2}, \tfrac{1}{2}\rangle\,, \qquad n = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |N; \tfrac{1}{2}, -\tfrac{1}{2}\rangle\,, \\ \bar{p} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\bar{N}; \tfrac{\bar{1}}{2}, -\tfrac{1}{2}\rangle\,, \qquad \bar{n} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -|\bar{N}; \tfrac{\bar{1}}{2}, \tfrac{1}{2}\rangle\,. \end{aligned} \tag{5.5}$$

This can also be expressed as $I_3(p,n) = (+\frac{1}{2}, -\frac{1}{2})$ and $I_3(\bar{n},\bar{p}) = (+\frac{1}{2}, -\frac{1}{2})$.

Consider next the three pions, (π^+, π^0, π^-). These particles have very similar masses, with $m_{\pi^\pm} = 139.57\,\text{MeV}$ and $m_{\pi^0} = 134.98\,\text{MeV}$, and are naturally classified as an isospin *triplet*, $I = 1$. If we decide to assign increasing I_3 to particles with increasing electric charge, as we did for the nucleon, then $I_3(\pi^+, \pi^0, \pi^-) = (+1, 0, -1)$. Similarly, one finds three Σ baryons, $(\Sigma^+, \Sigma^0, \Sigma^-$, with $m_{\Sigma^+} = 1.189\,\text{GeV}$, $m_{\Sigma^0} = 1.193\,\text{GeV}$ and $m_{\Sigma^-} = 1.19\,\text{GeV}$ ($\Sigma^\pm$ are not each other antiparticle!), naturally classified into an isotriplet, with $I_3(\Sigma^+, \Sigma^0, \Sigma^-) = (+1, 0, -1)$.

There are instead four kaons, the charged $K^\pm$, and the neutral K^0 and its antiparticle $\bar{K}^0$, with $m_{K^\pm} = 493.7\,\text{MeV}$ and $m_{K^0, \bar{K}^0} = 497.6\,\text{MeV}$. One may think that they form a quartet, but

[65] Locality requirements in quantum field theory force us to impose that particles and antiparticles transform in relatively complex-conjugate representations. This is consistent with the fact that particles and antiparticles have opposite internal quantum numbers: in particular, I_3 must anticommute with charge conjugation. More explicitly, to a set of particles and antiparticles with creation operators $b_i^\dagger$ and $d_i^\dagger$ correspond the fields $\psi_i \sim b_i + d_i^\dagger$ and their Hermitean conjugates. Under an isospin transformation $b_i^\dagger \to D_{ij} b_i^\dagger$ and so $b_i \to D_{ij}^* b_i$, and for the field to transform in a simple way we have to require $d_i^\dagger \to D_{ij}^* d_i^\dagger$, so that $\psi_i \to D_{ij}\psi_i$. Collecting all the ψ_i into a single multi-component field $\Psi = (\psi_1, \ldots, \psi_n)$ we have $\Psi = \sum_n \psi_i e_i$ with $(e_i)_j = \delta_{ij}$, and so the same representative in isospin space is associated with a particle and its antiparticle.

these states are further distinguished by another quantum number, their *strangeness*. Before the discovery of quarks, this was just another good quantum number one could consistently assign to particles, conserved in strong interaction processes, and compatible (commuting) with isospin.[66] For the kaons one has $S_{K^+,K^0} = 1$ and $S_{K^-,\bar{K}^0} = -1$, and so they are classified as two isodoublets, $I_3(K^+, K^0) = (+\frac{1}{2}, -\frac{1}{2})$ and $I_3(\bar{K}^0, K^-) = (+\frac{1}{2}, -\frac{1}{2})$. This looks like a case of *accidental degeneracy*, i.e., more degeneracy than one would expect based on symmetry alone. We will see that this is not so accidental after all. Notice that strangeness conservation is associated with a $\mathrm{U}(1)_S$ symmetry, namely symmetry under the phase transformations $e^{i\phi S}$. More generally, for any conserved charge Q there is a $U(1)_Q$ (possibly internal) symmetry under $e^{i\phi Q}$, that leaves the Hamiltonian unchanged. While the representation of the SU(2) group of isospin rotations $\mathrm{SU}(2)_I$ corresponding to kaons is reducible, the representation of $\mathrm{SU}(2)_I \times \mathrm{U}(1)_S$ is irreducible.

We can continue with the four Δ resonances, $(\Delta^{++}, \Delta^+, \Delta^0, \Delta^-)$, all with masses $m_\Delta = 1.232\,\mathrm{GeV}$. These are classified into a *quartet*, $I = \frac{3}{2}$, with I_3 assignment $I_3(\Delta^{++}, \Delta^+, \Delta^0, \Delta^-) = (+\frac{3}{2}, +\frac{1}{2}, -\frac{1}{2}, -\frac{3}{2})$. In this case there is in fact no other hadronic quantum number to distinguish the four states (electric charge is unknown to the strong interactions).

Finally, isosinglets are also found in the hadron spectrum, e.g., the η meson and the Λ baryon, with no partners with similar mass and the same strangeness.

Gell-Mann–Nishijima formula Two empirical facts were observed concerning multiplets. The first is that no two states with the same electric charge belonged to the same multiplet. This allows us to assign an increasing value of I_3 to particles with increasing charge, as already mentioned above. Moreover, no gaps in electric charge where observed in multiplets, with particles showing charges from a minimal to a maximal value increasing in steps of 1. This means that I_3 and the electric charge Q are in a linear relation. As a matter of fact, the following empirical relation was found between Q, I_3 (assigned as explained above), baryon number B, and strangeness S, known as the *Gell-Mann–Nishijima formula*:

$$Q = I_3 + \frac{1}{2}(B + S) = I_3 + \frac{1}{2}Y\,. \tag{5.6}$$

For future utility, we have introduced the *hypercharge* $Y = B + S$. Since both B and S are conserved, so is Y. The U(1) symmetries associated with baryon number and strangeness conservation can then be traded for the U(1) symmetries associated with baryon number and hypercharge conservation, $U(1)_S \times U(1)_B = U(1)_Y \times U(1)_B$.

Origin of isospin symmetry In hindsight, knowing of the existence of quarks and of QCD, we can trace isospin symmetry back to the fact that strong interactions are blind to the flavour of quarks, except for the difference in their masses. If quark masses were identical, then we could "rotate" flavours into each other without any physical effect. The two lightest quarks,

[66] "Strangeness" was literally associated with the strange behaviour of certain particles. Strange particles were produced on short time scales, typical of the strong interactions, but decayed on long time scales, typical of the weak interactions. It was then natural to assume that strong and weak interactions were responsible for their production and decay, respectively. In the production process only certain *pairs* of strange particles always appeared, signaling the existence of a new quantum number conserved by the strong interactions: it was indeed possible to consistently assign an integer number called *strangeness* to each particle, in such a way that it was conserved in strong processes. On the other hand, strangeness was not conserved in weak processes.

the up and the down, happen to have masses whose difference is much smaller than the typical hadronic scale $\Lambda \sim 100$ MeV $\div$ 1 GeV, and so they effectively degenerate in mass to a good extent. QCD has then an approximate but very accurate symmetry under SU(2) rotations in the space of the up and down quarks, which (leaving aside electromagnetic and weak effects) is broken only by the small difference between their masses. In modern terms, we treat u and d and their antiparticles $\bar{u}$ and $\bar{d}$ as two isodoublets, assigning them representative vectors as we did for the proton and neutron in Eq. (5.5), i.e.,

$$\begin{aligned} u &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |q; \tfrac{1}{2}, \tfrac{1}{2}\rangle \,, \qquad & d &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |q; \tfrac{1}{2}, -\tfrac{1}{2}\rangle \,, \\ \bar{u} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\bar{q}; \tfrac{\bar{1}}{2}, -\tfrac{1}{2}\rangle \,, \qquad & \bar{d} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -|\bar{q}; \tfrac{\bar{1}}{2}, \tfrac{1}{2}\rangle \,, \end{aligned} \tag{5.7}$$

where q and $\bar{q}$ denote the (fictitious) light-quark and light-antiquark particles, of which u, d and $\bar{u}, \bar{d}$ are just two of all the possible states. We then notice that the Hamiltonian of strong interactions is invariant under unitary rotations in (u, d) space, up to a small symmetry breaking term proportional to $(m_d - m_u)/\Lambda$. This explains the origin and the goodness of isospin symmetry. The strange quark s can of course be included in the picture, although on a different footing than u and d since it has a considerably different mass: one simply leaves it alone, as it is trivially unaffected by isospin transformations. In mathematical terms this means that s is an isosinglet, $s = |s; 0, 0\rangle$. The same is done with the heavier quarks as well; in general, all elementary particles different from u and d are assigned $I = 0$.

Isospin multiplets from quarks Using the rules for the composition of SU(2) representations, and recalling that mesons are quark-antiquark states, one finds that mesons should come in singlets and doublets, for states made only of u and d; and singlets and doublets for states involving a strange quark or antiquark. Baryons instead are quark-quark-quark states, and so the non-strange ones should come in doublets and quartets; the singly-strange in singlets and triplets; the doubly-strange in doublets; and the triply-strange in singlets. This follows from the rules

$$\begin{aligned} q\bar{q}: &\qquad \frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1 \,, \\ s\bar{q}: &\qquad 0 \otimes \frac{1}{2} = \frac{1}{2} \,, \\ q\bar{s}: &\qquad \frac{1}{2} \otimes 0 = \frac{1}{2} \,, \end{aligned} \tag{5.8}$$

that apply to mesons, and

$$\begin{aligned} qqq: &\qquad \frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = (0 \oplus 1) \otimes \frac{1}{2} = \frac{1}{2} \oplus \frac{1}{2} \oplus \frac{3}{2} \\ sqq: &\qquad 0 \otimes \frac{1}{2} \otimes \frac{1}{2} = 0 \otimes (0 \oplus 1) = 0 \oplus 1 \,, \\ ssq: &\qquad 0 \otimes 0 \otimes \frac{1}{2} = \frac{1}{2} \,, \\ sss: &\qquad 0 \otimes 0 \otimes 0 = 0 \,, \end{aligned} \tag{5.9}$$

that apply to baryons.

To see in detail how this works, consider the mesons containing only u and d. From Eq. (5.8) we see that we get an isosinglet and an isotriplet. An isosinglet can also be obtained as $s\bar{s}$, and the two states may mix; the resulting particles are the mesons η and η'. The isotriplet, on the other hand, corresponds to the pions:

$$\begin{aligned} -u\bar{d} &= |q;\tfrac{1}{2}\,\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,\tfrac{1}{2}\rangle &&= |1\,1\rangle &&= \pi^+\,, \\ \tfrac{1}{\sqrt{2}}(u\bar{u}-d\bar{d}) &= \tfrac{1}{\sqrt{2}}\left(|q;\tfrac{1}{2}\,\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,-\tfrac{1}{2}\rangle + |q;\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,\tfrac{1}{2}\rangle\right) &&= |1\,0\rangle &&= \pi^0\,, \\ d\bar{u} &= |q;\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &&= |1\,-1\rangle &&= \pi^-\,. \end{aligned} \tag{5.10}$$

Kaons are obtained even more simply, as it is clear from Eq. (5.8). One finds straightforwardly

$$\begin{aligned} u\bar{s} &= |q;\tfrac{1}{2}\,\tfrac{1}{2}\rangle \otimes |\bar{s}\rangle &&= |\tfrac{1}{2}\,\tfrac{1}{2}\rangle &&= K^+\,, \\ d\bar{s} &= |q;\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \otimes |\bar{s}\rangle &&= |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &&= K^0\,, \\ s\bar{u} &= |s\rangle \otimes |\bar{q};\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &&= |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &&= K^-\,, \\ -s\bar{d} &= |s\rangle \otimes |\bar{q};\tfrac{1}{2}\,\tfrac{1}{2}\rangle &&= |\tfrac{1}{2}\,\tfrac{1}{2}\rangle &&= \bar{K}^0\,. \end{aligned} \tag{5.11}$$

with the two multiplets distinguished by strangeness.

Quarks also allow us to understand how the Gell-Mann–Nishijima formula comes along: if the charges Q, B, and S of the three quarks u, d, s, satisfy it, then any of the plethora of hadrons built out of them and their antiquarks will automatically satisfy it. Assigning charges that fulfill the Gell-Mann–Nishijima formula is indeed possible: this will be discussed in detail later [see Section 5.2.5, in particular Eq. (5.77)].

To see that such an assignment is possible, consider the proton and the neutron. The relevant composition is in the first line of Eq. (5.9), and taking in particular the $\frac{1}{2}$ appearing as $\frac{1}{2}\otimes\frac{1}{2}\otimes\frac{1}{2}\to 0\otimes\frac{1}{2}$, we have

$$p \sim \frac{ud-du}{\sqrt{2}}u \sim uud\,, \qquad n \sim \frac{ud-du}{\sqrt{2}}d \sim udd\,. \tag{5.12}$$

Proton and neutron have baryon number $B_p = B_n = 1$, strangeness $S_p = S_n = 0$, and electric charge $Q_p = 1$ and $Q_n = 0$. Given the quark content of Eq. (5.12), one finds the equations

$$\begin{aligned} &1 = 2q_u + q_d\,, \qquad 1 = 2b_u + b_d\,, \qquad 0 = 2s_u + s_d\,, \\ &0 = q_u + 2q_d \qquad 1 = b_u + 2b_d\,, \qquad 0 = 2s_u + s_d\,, \end{aligned} \tag{5.13}$$

that are solved by

$$\begin{aligned} &q_u = \frac{2}{3}\,, \qquad b_u = \frac{1}{3}\,, \qquad s_u = 0\,, \\ &q_d = -\frac{1}{3}\,, \qquad b_d = \frac{1}{3}\,, \qquad s_d = 0\,. \end{aligned} \tag{5.14}$$

To fix the charges of the strange quark, we use the $\bar{K}^0$, with vanishing electric charge and baryon number and strangeness -1. Using Eq. (5.11) we find

$$q_s = q_d = -\frac{1}{3}\,, \qquad b_s = b_d = \frac{1}{3}\,, \qquad s_s = -1\,. \tag{5.15}$$

Intermezzo: two-dimensional harmonic oscillator One might wonder how come that starting from asking for invariance under exchange of proton and neutron we ended up with the bigger SU(2) symmetry. To see that this is in fact a natural implementation of the initial requirement, we discuss now a simple toy model, namely the two-dimensional harmonic oscillator. In terms of creation and annihilation operators, and up to an irrelevant additive constant, the Hamiltonian of this system reads

$$H_{ho} = \hbar\omega(a_1^\dagger a_1 + a_2^\dagger a_2)\,. \tag{5.16}$$

The creation and annihilation operators satisfy the commutation relations

$$[a_i, a_j] = [a_i^\dagger, a_j^\dagger] = 0\,, \qquad [a_i, a_j^\dagger] = \delta_{ij}\,. \tag{5.17}$$

The creation and annihilation operators can be interpreted as follows: $a_1^\dagger$ and a_1 create and destroy a static "bosonic proton", while $a_2^\dagger$ and a_2 create and destroy a static "bosonic neutron". No interaction takes place between these particles. The general eigenstate of H_{ho} reads

$$|n_1, n_2\rangle = \frac{1}{\sqrt{n_1!}}\frac{1}{\sqrt{n_2!}}(a_1^\dagger)^{n_1}(a_2^\dagger)^{n_2}|0\rangle\,, \tag{5.18}$$

with $|0\rangle$ the ground state $a_{1,2}|0\rangle = 0$. The total energy of the system in a given eigenstate is then obtained summing the number of protons, n_1, to the number of neutrons, n_2, times their (equal) mass $\hbar\omega$: $E_{n_1,n_2} = \mathcal{E}_{n_1+n_2} = \hbar\omega(n_1+n_2)$. It is straightforward to establish that the degeneracy of the energy level $\mathcal{E}_n$ is $n+1$. If we apply the operators $I_- = a_2^\dagger a_1$ and $I_+ = a_1^\dagger a_2$ to a state $|n_1, n_2\rangle$ we obtain

$$a_2^\dagger a_1|n_1, n_2\rangle \propto |n_1 - 1, n_2 + 1\rangle\,, \qquad a_1^\dagger a_2|n_1, n_2\rangle \propto |n_1 + 1, n_2 - 1\rangle\,, \tag{5.19}$$

i.e., these operators respectively replace a proton with a neutron and a neutron with a proton. It is evident that $I_+ = I_-^\dagger$, and one can verify explicitly that $[I_\pm, H_{ho}] = 0$. Moreover, defining $I_3 = \frac{1}{2}(a_1^\dagger a_1 - a_2^\dagger a_2)$, one finds that also $[I_3, H_{ho}] = 0$. This is an obvious consequence of the fact that $2I_3 = [I_+, I_-]$, as can be directly verified. Notice now that the Hamiltonian is manifestly invariant under unitary rotations of the creation and annihilation operators, i.e., $a_i \to U_{ij}a_j$ with $U \in \mathrm{SU}(2)$. It can be verified explicitly that I_3, $I_1 = (I_+ + I_-)/2$ and $I_2 = (I_+ - I_-)/(2i)$ are indeed the generators of these transformations, and satisfy $[I_i, I_j] = i\varepsilon_{ijk}I_k$. The proton-neutron exchange operators are then naturally part of an SU(2) algebra. One can go on and show that $I^2 = I_1^2 + I_2^2 + I_3^2 = \frac{H_{ho}}{2\hbar\omega}(\frac{H_{ho}}{2\hbar\omega} + 1)$. Eigenstates of energy $\mathcal{E}_n = \hbar\omega n$ form therefore a multiplet of isospin $I = \frac{n}{2}$, which leads to degeneracy $2I + 1 = n + 1$. Incidentally, this explains the accidental degeneracy of the eigenstates of the two-dimensional harmonic oscillator.

Dynamical consequences of isospin invariance Besides static properties, isospin invariance has important consequences for dynamical processes as well. Consider for example the η meson, $m_\eta = 547$ MeV. This is a neutral pseudoscalar meson ($J_\eta = 0$, $\eta_\eta = -1$) like the π^0, but with $I = 0$. As such, it can decay electromagnetically in two photons. Its mass is sufficient also for strong decays into two and three pions, but the two-pion decay process is forbidden by parity: $J = 0$ in the final state requires $\ell = 0$, so that it would have positive rather than negative parity. The three-pion process is instead forbidden by isospin and charge-conjugation symmetry. In fact, electric charge conservation restricts the possible three-pion final states to $\pi^0\pi^0\pi^0$ and $\pi^+\pi^-\pi^0$. One can show that under the isospin rotation $R_2 = e^{i\pi I_2}$ the following relation holds,

$$R_2|I\, I_3\rangle = (-1)^{I-I_3}|I\ -I_3\rangle\,. \tag{5.20}$$

Moreover, both the eta and neutral pion have charge-conjugation parity $\xi_\eta = \xi_{\pi^0} = 1$. Defining the *G-parity* transformation as $G \equiv CR_2$, one then finds

$$\begin{aligned} G|\eta\rangle &= C|\eta\rangle = |\eta\rangle\,, \\ G|\pi^+\pi^-\pi^0\rangle &= C(-1)^{1-1}(-1)^{1+1}(-1)^{1-0}|\pi^-\pi^+\pi^0\rangle = -|\pi^+\pi^-\pi^0\rangle\,, \\ G|\pi^0\pi^0\pi^0\rangle &= C(-1)^3|\pi^0\pi^0\pi^0\rangle = -|\pi^0\pi^0\pi^0\rangle\,, \end{aligned} \tag{5.21}$$

so that the G-parity of the eta and of the three-pion state differ.

Isospin conservation implies also quantitative relations between the probabilities of different scattering processes taking place. As we will see later, the probability that two particles scatter and produce some prescribed final state is given by the absolute value square $|\mathcal{M}_{i\to f}|$ of the *scattering amplitude*, which in turn is the matrix element of the scattering operator S (or *S-matrix*) between the free-particle states $|\phi_i\rangle$ and $|\phi_f\rangle$ corresponding to the initial and final states of the process, $\mathcal{M}_{i\to f} = \langle\phi_f|S|\phi_i\rangle$. All that we need to know at this stage is that if both the free and the interaction part of the Hamiltonian are invariant under a symmetry transformation, so will be the S-matrix, i.e., if for some symmetry transformation M we have $[M, H_0] = [M, H] = 0$ then $[M, S] = 0$. Let us consider the case of strong interactions and isospin invariance, and focus on nucleon-pion scattering processes. To this end, it is convenient to decompose the nucleon-pion states in pure isospin components. From the composition rule $\frac{1}{2}\otimes 1 = \frac{1}{2}\oplus\frac{3}{2}$ we find that this states contain a pure $I = \frac{1}{2}$ and a pure $I = \frac{3}{2}$ part. A simple calculation using the lowering operators I_- and the relation $\vec{I}^2 = I_+I_- + I_3(I_3 - 1)$ shows that

$$\begin{aligned} |\tfrac{3}{2}\,\tfrac{3}{2}\rangle &= |\tfrac{1}{2}\,\tfrac{1}{2}\rangle|1\,1\rangle &&= |p\pi^+\rangle\,, \\ \sqrt{3}|\tfrac{3}{2}\,\tfrac{1}{2}\rangle &= |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|1\,1\rangle + \sqrt{2}|\tfrac{1}{2}\,\tfrac{1}{2}\rangle|1\,0\rangle &&= |n\pi^+\rangle + \sqrt{2}|p\pi^0\rangle\,, \\ \sqrt{3}|\tfrac{3}{2}\,-\tfrac{1}{2}\rangle &= \sqrt{2}|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|1\,0\rangle + |\tfrac{1}{2}\,\tfrac{1}{2}\rangle|1\,-1\rangle &&= \sqrt{2}|n\pi^0\rangle + |p\pi^-\rangle\,, \\ |\tfrac{3}{2}\,-\tfrac{3}{2}\rangle &= |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|1\,-1\rangle &&= |n\pi^-\rangle\,, \\ \sqrt{3}|\tfrac{1}{2}\,\tfrac{1}{2}\rangle &= \sqrt{2}|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|1\,1\rangle - |\tfrac{1}{2}\,\tfrac{1}{2}\rangle|1\,0\rangle &&= \sqrt{2}|n\pi^+\rangle - |p\pi^0\rangle\,, \\ \sqrt{3}|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &= |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|1\,0\rangle - \sqrt{2}|\tfrac{1}{2}\,\tfrac{1}{2}\rangle|1\,-1\rangle &&= |n\pi^0\rangle - \sqrt{2}|p\pi^-\rangle\,. \end{aligned} \tag{5.22}$$

These relations can be inverted to achieve the desired decomposition. Here we will be concerned with the processes

$$p\;\pi^+ \to p\;\pi^+\,, \qquad p\;\pi^- \to p\;\pi^-\,, \qquad p\;\pi^- \to n\;\pi^0\,, \tag{5.23}$$

so we will need the following results,

$$\begin{aligned} |p\pi^+\rangle &= |\tfrac{3}{2}\,\tfrac{3}{2}\rangle\,, \\ |p\pi^-\rangle &= \tfrac{1}{\sqrt{3}}\left(|\tfrac{3}{2}\,-\tfrac{1}{2}\rangle - \sqrt{2}|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle\right)\,, \\ |n\pi^0\rangle &= \tfrac{1}{\sqrt{3}}\left(\sqrt{2}|\tfrac{3}{2}\,-\tfrac{1}{2}\rangle + |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle\right)\,. \end{aligned} \tag{5.24}$$

Exploiting invariance one can prove that

$$\begin{aligned} \langle I'I'_3|S|II_3\rangle &= \delta_{I'I}\delta_{I'_3I_3}\langle II_3|S|II_3\rangle\,, \\ \langle II_3|S|II_3\rangle &= \langle II'_3|S|II'_3\rangle \equiv \mathcal{M}_I\,. \end{aligned} \tag{5.25}$$

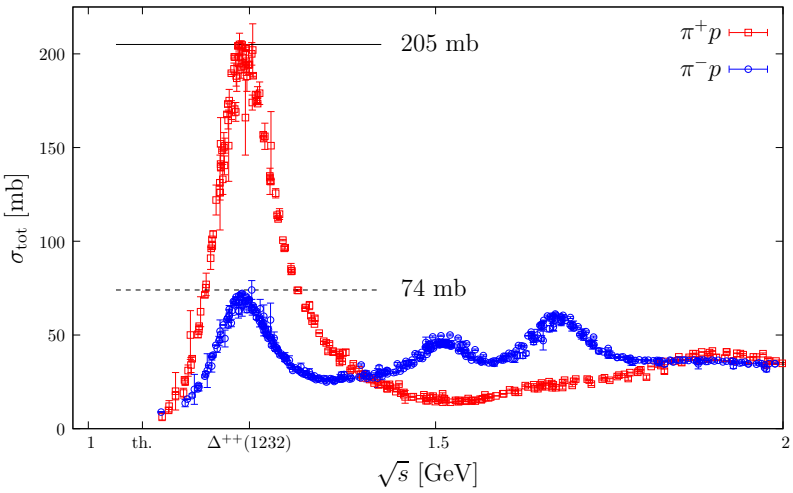

Figure 19: Total cross sections for π^+p and π^-p scattering.

Both these results follow from $[\vec{I}, S] = 0$: in particular, $[\vec{I}^2, S] = 0$ and $[I_3, S] = 0$ imply that the initial and final state must have the same eigenvalues of $\vec{I}^2$ and I_3 for the matrix element not to vanish. The second result can be proved explicitly using $[I_\pm, S] = 0$, but is in fact a particular case of the more general Wigner-Eckart theorem. We then find for the nucleon-pion scattering amplitudes

$$\begin{aligned}
\mathcal{M}_{p\pi^+\to p\pi^+} &= \langle p\pi^+|S|p\pi^+\rangle = \langle \tfrac{3}{2}\ \tfrac{3}{2}|S|\tfrac{3}{2}\ \tfrac{3}{2}\rangle = \mathcal{M}_{\frac{3}{2}}\,, \\
\mathcal{M}_{p\pi^-\to p\pi^-} &= \langle p\pi^-|S|p\pi^-\rangle = \tfrac{1}{3}\left(\langle \tfrac{3}{2}\ -\tfrac{1}{2}|S|\tfrac{3}{2}\ -\tfrac{1}{2}\rangle + 2\langle \tfrac{1}{2}\ -\tfrac{1}{2}|S|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle\right) = \tfrac{1}{3}\left(\mathcal{M}_{\frac{3}{2}} + 2\mathcal{M}_{\frac{1}{2}}\right)\,, \\
\mathcal{M}_{p\pi^-\to n\pi^0} &= \langle n\pi^0|S|p\pi^-\rangle = \tfrac{\sqrt{2}}{3}\left(\langle \tfrac{3}{2}\ -\tfrac{1}{2}|S|\tfrac{3}{2}\ -\tfrac{1}{2}\rangle - \langle \tfrac{1}{2}\ -\tfrac{1}{2}|S|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle\right) = \tfrac{\sqrt{2}}{3}\left(\mathcal{M}_{\frac{3}{2}} - \mathcal{M}_{\frac{1}{2}}\right)\,.
\end{aligned} \tag{5.26}$$

Notice that one also has $\mathcal{M}_{p\pi^+\to n\pi^0} = \mathcal{M}_{n\pi^0\to p\pi^+}$. Other relations between amplitudes can be derived along the lines above. It is an experimentally known fact that at a centre-of-mass energy of $\sqrt{s} = 1.232$ GeV $= m_\Delta$ the $p\pi^+$ scattering process shows a peak in its cross section, which is proportional to the probability of the process (see Section 2). This peak corresponds precisely to the Δ^{++} resonance: at this energy the scattering process proceeds through the formation of this unstable particle and its subsequent decay. Such a particle is a member of an isospin quartet, with $I = \frac{3}{2}$, so it is expected that for energies near m_Δ the amplitude $\mathcal{M}_{\frac{3}{2}}$ will dominate over $\mathcal{M}_{\frac{1}{2}}$: one can imagine the S-matrix element is approximately the product of the amplitude to create the Δ times the amplitude for its decay, which involve only the $I = \frac{3}{2}$ components of the scattering states. At $\sqrt{s} \approx m_\Delta$ we will then have that At those energies we will have that the cross sections for the processes in Eq. (5.23) satisfy[67]

$$\frac{\sigma_{p\pi^+\to p\pi^+}}{\sigma_{p\pi^-\to p\pi^-}} \simeq 9\,, \qquad \frac{\sigma_{p\pi^+\to p\pi^+}}{\sigma_{p\pi^-\to n\pi^0}} \simeq \frac{9}{2}\,. \tag{5.27}$$

[67]The proportionality factors between cross section and $|\mathcal{M}_{i\to f}|^2$ depend on the mass and spin of the particles and on $\sqrt{s}$, so they are the same for all the processes and cancel out in ratios.

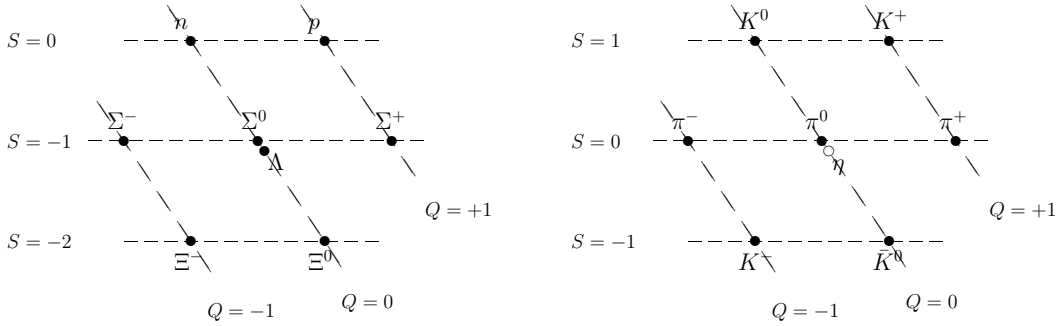

Figure 20: Lightest spin $\frac{1}{2}$ baryons (left) and lightest pseudoscalar mesons (right) known at the end of the '50s (black circles) organised in hexagonal patterns. Horizontal lines correspond to constant strangeness, diagonal lines correspond to constant electric charge. The η meson is shown as well with an empty circle.

At this energy, the elastic channel is essentially the only scattering channel available in a $p\pi^+$ scattering process (there is enough energy to produce an extra neutral pion, but the available phase space is very small and the corresponding cross section is negligible), while for $p\pi^-$ there are the elastic channel and the inelastic process $p\pi^- \to n\pi^0$ (and $p\pi^- \to p\pi^-\pi^0$, $n\pi^0\pi^0$ as well, but suppressed). We can then obtain the following relation for total cross sections,

$$\frac{\sigma_{p\pi^+\mathrm{tot}}}{\sigma_{p\pi^-\mathrm{tot}}} = \frac{\sigma_{p\pi^+\to p\pi^+}}{\sigma_{p\pi^-\to p\pi^-} + \sigma_{p\pi^-\to n\pi^0}} \underset{\sqrt{s}=m_\Delta}{\simeq} \frac{1}{\frac{1}{9}+\frac{2}{9}} = 3\,. \tag{5.28}$$

This relation is well verified experimentally (see Fig. 19).

A similar but easier calculation can be done to relate the three inelastic processes

$$pp \to d\pi^+\,, \quad pn \to d\pi^0\,, \quad nn \to d\pi^-\,. \tag{5.29}$$

The deuteron d has $I = 0$, so the final states are pure $I = 1$ states. The initial states have in general both $I = 0$ and $I = 1$, since $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$, but only the $I = 1$ component will contribute to the scattering amplitude. More precisely, the pp and nn states are pure $I = 1$, while for the pn state we have

$$|pn\rangle = \frac{1}{\sqrt{2}}(|10\rangle + |00\rangle)\,. \tag{5.30}$$

One then concludes that $P_{pp\to d\pi^+} = P_{nn\to d\pi^-} = 2P_{pn\to d\pi^0}$.

5.2 The quark model

Knowing as we do now the microscopic theory of strong interactions, i.e., QCD, it would be easy to envisage larger approximate symmetry groups relating different flavours of quarks, broken more and more badly by the increasing differences in their mass. Historically, things went the opposite way, with a larger approximate symmetry discovered first, and its origin and the origin of its breaking understood later in terms of fundamental constituents bound together to form the hadrons observed in experiments.

By the end of the '50s, the known hadrons had grown into a "zoo", comprising the two big families of baryons and mesons, which could be further classified in isospin multiplets (see

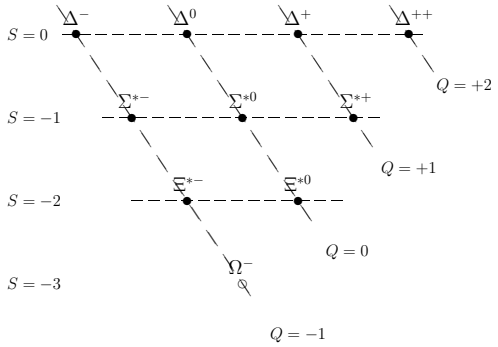

Figure 21: Spin-$\frac{3}{2}$ baryon resonances known at the end of the '50s (black circles) organised in a triangular pattern. Horizontal lines correspond to constant strangeness, diagonal lines correspond to constant electric charge. The Ω baryon is shown as well with an empty circle.

above) and according to their strangeness. In fact, the isospin quantum numbers I, I_3, the baryon number B, and the strangeness S allowed a full classification of the known hadrons, entirely characterising their states together with four-momentum and spin (total and one of its components). Baryon number and strangeness could be equivalently traded for baryon number and hypercharge $Y = B+S$. The hadron "zoo" seemed to lack an organising principle, and some thought that there was not any: this was the idea of "nuclear democracy", in which hadrons are thought to be somehow all elementary and composite at the same time.[68] However, if one looked carefully enough certain patterns could be found, hinting at the possible existence of an organising principle in the form of an underlying approximate symmetry.

If one plotted the eight lightest, spin-$\frac{1}{2}$ baryons in the I_3, S (isospin-strangeness) plane, they would fit nicely in a hexagonal array (see Fig. 20). These baryons fit also into isospin multiplets, with small (permille) mass splittings within each multiplet. Mass splittings between baryons with different strangeness were larger, but showed an interesting regularity: a decrease of one unit in strangeness corresponded to a change in mass of the order of 150 MeV.

If one plotted instead the nine known spin-$\frac{3}{2}$ baryon resonances, one would find an almost-triangular array (see Fig. 21), with a single isospin multiplet for each value of S, and the same regularity concerning the mass splittings: with a decrease in S comes an extra 150 MeV of mass.

If one plotted the seven lightest pseudoscalar mesons one would find a hexagonal array similar to that of the light baryons, with a "missing meson" in the centre, and shifted by one unit in the direction of strangeness (see Fig. 20). Using instead the hypercharge $Y = B + S$, the two patterns would precisely overlap (see Fig. 22). Hypercharge seems therefore better suited than strangeness for a general classification of hadrons. Concerning mass splittings, they are still small within isospin multiplets, but they are rather pronounced between multiplets: being of the order of a few hundreds MeV, they are comparable with the meson masses.

It is typical of the human mind to look for explanations whenever some regularity appears in Nature. Here a simple explanation would be the existence of an approximate symmetry extending the $\mathrm{SU}(2)_I$ isospin symmetry and the $\mathrm{U}(1)_Y$ symmetry associated with hypercharge. The existence of the symmetry would explain the patterns through its irreducible representations; its breaking would explain the mass differences. More precisely, the approximate degeneracy could be explained if the strong Hamiltonian H_{strong} were the sum of some exactly symmetric Hamil-

[68] This was the principle behind the "bootstrap model" of G. Chew.

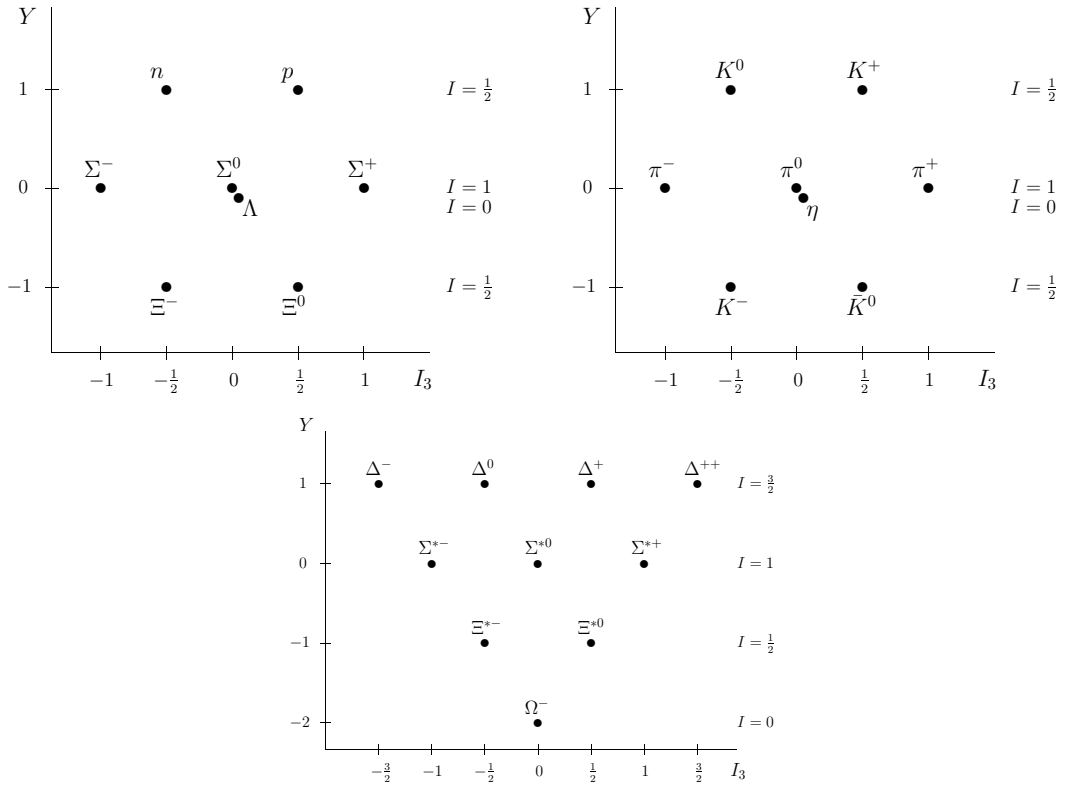

Figure 22: Baryons and mesons as in Figs. 20 and 21 but in the (I_3, Y) plane. Total isospin assignments are also shown.

tonian, with degenerate multiplets of eigenstates corresponding to irreducible representations of some continuous internal symmetry group, and a symmetry-breaking term, $H_{\text{strong}} = H_0 + H_I$. The symmetric Hamiltonian H_0 would commute with the symmetry generators $\mathcal{O}_a$, $[H_0, \mathcal{O}_a] = 0$, and so degenerate multiplets would emerge, exactly as in the case of isospin symmetry.

If symmetry is the reason behind the hadron multiplets, then the symmetry group should be such that it admits irreducible representations able to accomodate the experimentally observed hadronic multiplets. A reasonable set of assumptions reduces the quest to internal symmetry groups (i.e., that commute with the Poincaré group of translations and Lorentz transformations) that are compact Lie groups (see Ref. [10]). Being continuous groups, their representation on the Hilbert space of hadrons must be unitary. Furthermore, we already know part of the sought-after symmetry group: since we look for an extension of the isospin-hypercharge-baryon symmetry, whatever the bigger symmetry group G_{tot} is, it must contain $G_{\text{tot}} \supset \mathrm{SU}(2)_I \times \mathrm{U}(1)_Y \times \mathrm{U}(1)_B$. Since there is no (not even approximate) degeneracy observed between baryons and mesons, we do not expect the group G_{tot} to mix states different baryon numbers, and so $\mathrm{U}(1)_B$ should remain as an extra factor, $G_{\text{tot}} = G \times \mathrm{U}(1)_B$, and $G \supset \mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$. The representations of G should also reproduce the observed isospin multiplets within the bigger hadronic multiplets, together with the corresponding values of hypercharge. In particular, the baryon octet decomposes into two isospin doublets, one with hypercharge 1 (the nucleons) and one with hypercharge -1 (the Ξs), an isospin triplet with hypercharge 0 (the Σs), and an isospin singlet with hypercharge 0 (the Λ). Whatever G is, it should admit an eight dimensional representation made up of these

isospin multiplets; hoepfully, it would also provide the seven-dimensional and nine-dimensional multiplets of mesons and baryonic resonances.

The simplest group that satisfies the requirements above, leading to an eight-dimensional representation with the desired decomposition into isospin-hypercharge multiplets (and that does not contradict other experimental facts: see again Ref. [10]), is the group SU(3). We first discuss this group and its representations in some detail, and then see how it can be applied to the problem of the classification of hadrons.

5.2.1 The group SU(3) and its algebra

The group SU(3) is the group of unitary unimodular (i.e., with unit determinant) 3×3 complex matrices, i.e., $\mathrm{SU}(3) = \{U \in M_{3,3}(\mathbb{C}) \mid U^\dagger U = \mathbf{1}\,,\ \det U = 1\}$. In general, any $n \times n$ unitary matrix can be written as $U = e^{iA}$ with A Hermitian, since $\mathbf{1} = e^{iA}e^{-iA^\dagger} \Rightarrow A = A^\dagger$. The unimodularity condition $\det U = e^{i\mathrm{tr}\,A} = 1$ further requires $\mathrm{tr}\,A = 0$.[69] The space of Hermitian traceless matrices is an 8-dimensional real vector space, and any such matrix can be written as a linear combination with real coefficients, $A = \sum_a \alpha_a t^a$, of suitable linearly-independent basis elements $\{t^a\}_{a=1,\ldots,8}$.

Writing our matrices in the form $U(\alpha) = e^{i\alpha_a t^a}$ shows us that SU(3) is an 8-dimensional Lie group, with corresponding Lie algebra the space of Hermitian traceless matrices equipped with the matrix commutator. In fact, since $-i[A, B]$ is still a Hermitian traceless matrix, one can write it as a linear combination with real coefficients of the basis matrices t^a, i.e., the generators of the group, which obey the commutation relations

$$[t^a, t^b] = i f_{abc} t^c\,, \tag{5.31}$$

with real f_{abc} called the *structure constants* of the group. These are obviously antisymmetric in the first pair of indices, $f_{bac} = -f_{abc}$. The generators are usually chosen with the normalisation[70]

$$\mathrm{tr}\,t^a t^b = \tfrac{1}{2}\delta^{ab}\,. \tag{5.32}$$

Using Eq. (5.32), one finds

$$f_{abc} = -2i\,\mathrm{tr}\,[t_a, t_b]t_c\,, \tag{5.33}$$

which shows that with this choice the structure constants are cyclic, $f_{abc} = f_{cab} = f_{bca}$, and so antisymmetric under exchange of any pair of indices. Since the commutators of any three matrices A, B, C satisfy the *Jacobi identity*,

$$[[A, B], C] + [[C, A], B] + [[B, C], A] = 0\,, \tag{5.34}$$

[69] More precisely, since $[U, U^\dagger] = 0$, any $U \in \mathrm{U}(n)$ can be diagonalised by a unitary transformation, and it is easy to see that its eigenvalues are phase factors $e^{i\phi_a}$, $\phi_a \in \mathbb{R}$. Denoting $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_N)$, one has $U = V^\dagger e^{i\Phi} V = e^{iV^\dagger \Phi V} = e^{iA}$, with $A = V^\dagger \Phi V = A^\dagger$. Using the relation $\det e^{iA} = e^{i\mathrm{tr}\,A}$, we find in general $\mathrm{tr}\,A = 2\pi k$ with $k \in \mathbb{Z}$, so $A = A_{\mathrm{traceless}} + \frac{2\pi k}{N}$. But

$$e^{i\frac{2\pi k}{N}}\mathbf{1} = e^{i\frac{2\pi k}{N}\mathrm{diag}(1,1,\ldots,1,1)} = e^{i\frac{2\pi k}{N}\mathrm{diag}(1,1,\ldots,1,1-N)} = e^{i\frac{2\pi k}{N}z}\,,$$

with $z = \mathrm{diag}(1, 1, \ldots, 1, 1 - N)$ a traceless matrix, and so we can always express U as $U = e^{i(A_{\mathrm{traceless}} + \frac{2\pi k}{N}z)} = e^{iA'_{\mathrm{traceless}}}$.

[70] It is always possible to choose the generators such that Eq. (5.32) is satisfied. Since the matrix $\mathrm{tr}\,t^a t^b$ is a positive-definite real symmetric matrix, it can always be diagonalised by an orthogonal change of basis $t^a \to \tilde{t}^a = O^a t^b$, and furthermore the resulting positive diagonal entries can be set to any other real value by a rescaling of $\tilde{t}^a$.

one finds

$$[[t^a,t^b],t^c]+[[t^c,t^a],t^b]+[[t^b,t^c],t^a]=0 \Longrightarrow f_{bcm}f_{amn}+f_{abm}f_{cmn}+f_{cam}f_{bmn}=0\,. \tag{5.35}$$

Even after imposing the normalisation condition there is still a large freedom in choosing the t^a, as these are just basis elements of a vector space. A general change of basis of the algebra leads to an equally good set of generators, but in general with different structure constants and normalisation.

A convenient choice of generators is guided by our intention to eventually break the symmetry to the smaller group $\mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$: we want then three of the generators to be isospin generators, reproducing the algebra of SU(2), and a fourth one to be the hypercharge generator, commuting with the isospin ones. We then take $t^a = \frac{1}{2}\lambda^a$, with λ^a the *Gell-Mann matrices*

$$\begin{aligned}
\lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
\lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \qquad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\
\lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
\end{aligned} \tag{5.36}$$

One can verify explicitly that with this choice the structure constants satisfy the normalisation condition[71]

$$f_{abc}f_{abd} = 3\delta_{cd}\,. \tag{5.37}$$

It is easy to identify $t^{1,2,3}$ as the generators $\vec{I}$ of the $\mathrm{SU}(2)_I$ isospin subgroup of SU(3), i.e., the SU(2) subgroup embedded in the top left corner of the SU(3) matrices; and t^8 as the generator Y of the $\mathrm{U}(1)_Y$ subgroup (up to a numerical factor). Indeed, we see without effort that

$$[t^a,t^b] = i\epsilon_{abc}t^c\,, \quad a,b,c=1,2,3\,, \qquad [t^8,t^a]=0\,, \quad a=1,2,3\,, \tag{5.38}$$

reproducing the commutation relations

$$[I_a,I_b] = i\epsilon_{abc}I_c\,, \quad a,b,c=1,2,3\,, \qquad [Y,I_a]=0\,, \quad a=1,2,3\,. \tag{5.39}$$

We then identify

$$t^a = I_a\,, \quad a=1,2,3, \qquad t^8 = \tfrac{\sqrt{3}}{2}Y\,. \tag{5.40}$$

The normalisation of Y will soon become clear, and has been chosen anticipating the result. With the choice Eq. (5.36), the canonical basis vectors $e^{(j)}$ of $\mathbb{C}^3$, $e_i^{(j)} = \delta_{ij}$, $j=1,2,3$, are simultaneous eigenvectors of I_3 and Y, which are the only two diagonal generators.

[71] The possibility to choose the generators such that the structure constants obey Eq. (5.37) (up to a positive factor) and are cyclic is not generic, but related to specific properties of the group, namely semisimplicity and compactness, which are features of SU(3) (and of any SU(n) in general).

The matrices $t^{4,5}$ and $\frac{1}{2}t^3 + \frac{\sqrt{3}}{2}t^8$ correspond to another SU(2) subgroup, the $\mathrm{SU}(2)_V$ subgroup, and $t^{6,7}$ and $-\frac{1}{2}t^3 + \frac{\sqrt{3}}{2}t^8$ correspond to a third SU(2) subgroup, the $\mathrm{SU}(2)_W$ subgroup. We then set

$$\begin{aligned} t^4 = V_1\,, \qquad t^5 = V_2\,, \qquad & \tfrac{1}{2}t^3 + \tfrac{\sqrt{3}}{2}t^8 = V_3 = \tfrac{1}{2}I_3 + \tfrac{3}{4}Y\,, \\ t^6 = W_1\,, \qquad t^7 = W_2\,, \qquad & -\tfrac{1}{2}t^3 + \tfrac{\sqrt{3}}{2}t^8 = W_3 = -\tfrac{1}{2}I_3 + \tfrac{3}{4}Y\,, \end{aligned} \tag{5.41}$$

and find immediately

$$[V_i, V_j] = i\epsilon_{ijk}V_k\,, \qquad [W_i, W_j] = i\epsilon_{ijk}W_k\,. \tag{5.42}$$

Notice that V_3 and W_3 are not new generators, but linear combinations of the existing ones. They are related to I_3 by the same permutation of the basis of $\mathbb{C}^3$ that relates $\vec{V}$ and $\vec{W}$ to $\vec{I}$. By the same reasoning, one sets

$$Y_V = I_3 - \tfrac{1}{2}Y\,, \qquad Y_W = -I_3 - \tfrac{1}{2}Y\,, \tag{5.43}$$

and shows straightforwardly that

$$[Y_V, V_i] = [Y_W, W_i] = 0\,. \tag{5.44}$$

In full analogy with the SU(2) case, we now define the ladder operators of our SU(2) subgroups, i.e., $I_\pm = I_1 \pm iI_2$, $V_\pm = V_1 \pm iV_2$ and $W_\pm = W_1 \pm iW_2$, which read explicitly

$$\begin{aligned} I_+ &= \frac{1}{2}(\lambda^1 + i\lambda^2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,, \qquad I_- = \frac{1}{2}(\lambda^1 - i\lambda^2) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,, \\ V_+ &= \frac{1}{2}(\lambda^4 + i\lambda^5) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,, \qquad V_- = \frac{1}{2}(\lambda^4 - i\lambda^5) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}\,, \\ W_+ &= \frac{1}{2}(\lambda^6 + i\lambda^7) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\,, \qquad W_- = \frac{1}{2}(\lambda^6 - i\lambda^7) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\,. \end{aligned} \tag{5.45}$$

It follows immediately that they satisfy the commutation relations

$$\begin{aligned} [I_3, I_\pm] = \pm I_\pm\,, \qquad & [V_3, V_\pm] = \pm V_\pm\,, \qquad [W_3, W_\pm] = \pm W_\pm\,, \\ [Y, I_\pm] = 0\,, \qquad & [Y_V, V_\pm] = 0\,, \qquad [Y_W, W_\pm] = 0\,. \end{aligned} \tag{5.46}$$

Combining now Eqs. (5.46) and (5.44) one finds

$$\begin{aligned} [I_3, I_\pm] = \pm I_\pm\,, \qquad & [Y, I_\pm] = 0\,, \\ [I_3, V_\pm] = \pm\tfrac{1}{2}V_\pm\,, \qquad & [Y, V_\pm] = \pm V_\pm\,, \\ [I_3, W_\pm] = \mp\tfrac{1}{2}W_\pm\,, \qquad & [Y, W_\pm] = \pm W_\pm\,. \end{aligned} \tag{5.47}$$

These relations are the SU(3) analogues of the commutation relations satisfied by the SU(2) ladder operators (which are contained here: they are the ones for $I_\pm$). Anticipating the discussion of the representations of the group, these relations show that the effect of $I_\pm$, $V_\pm$ and $W_\pm$ on a simultaneous eigenvector of I_3 and Y is to produce a new eigenvector with eigenvalues related in a simple way with the old ones. This also explains the normalisation of Y [see Eq. (5.40)]: the

hypercharges assigned to hadrons are integer numbers, so differing always by integer amounts. The $V_\pm$ and $W_\pm$ are meant to connect hadron states with hypercharge differing by one in the same multiplet, and so their effect on the simultaneous eigenvector of I_3 and Y has to be an increase or decrease of Y by one unit.

There are three more commutation relations that we can write down at once by simply recalling the results of SU(2), namely

$$[I_+, I_-] = 2I_3\,, \qquad [V_+, V_-] = 2V_3 = I_3 + \tfrac{3}{2}Y\,, \qquad [W_+, W_-] = 2W_3 = -I_3 + \tfrac{3}{2}Y\,. \tag{5.48}$$

From Eqs. (5.47) and (5.48) we can reconstruct part of the f_{abc} appearing in the commutation relations satisfied by the t^a, Eq. (5.31), corresponding to the choice Eq. (5.36).

The commutator as a linear operator on the algebra It will be very useful in the following to notice that Eq. (5.47) is nothing but a set of simultaneous eigenvalue equations for certain linear operators acting on the algebra of the group, i.e., the linear space spanned by the generators. While Y is most suited for our physical problem, from the mathematical point of view it is more convenient instead to use $\bar{Y} = \frac{\sqrt{3}}{2}Y$ (which is nothing but t^8) to express the commutation relations Eq. (5.47). Introducing the notation

$$\vec{H} = (I_3, \bar{Y})\,, \qquad E^{(1)}_\pm = I_\pm\,, \qquad E^{(2)}_\pm = V_\pm\,, \qquad E^{(3)}_\pm = W_\pm\,, \tag{5.49}$$

Eq. (5.47) can be expressed compactly as

$$[\vec{H}, E^{(j)}_\pm] = \pm\vec{\alpha}^{\,(j)} E^{(j)}_\pm\,, \tag{5.50}$$

where the *root vectors* $\vec{\alpha}^{\,(j)}$,

$$\vec{\alpha}^{\,(1)} = (1, 0)\,, \qquad \vec{\alpha}^{\,(2)} = (\tfrac{1}{2}, \tfrac{\sqrt{3}}{2})\,, \qquad \vec{\alpha}^{\,(3)} = (-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2})\,, \tag{5.51}$$

satisfy $(\vec{\alpha}^{\,(j)})^2 = 1$. The *root system* of SU(3) is shown in Fig. 23. As we already did for SU(2) [see Eq. (4.108)], we can look at the commutator $[A, X]$ for a fixed A as a linear function of X. For each A in the group algebra, this linear operator acting on the algebra is denoted ad_A,

$$\mathrm{ad}_A X \equiv [A, X]\,. \tag{5.52}$$

Notice that on the right-hand side X is treated as a matrix, whose commutator with A is obtained by the usual matrix operations, while on the left-hand side it is treated as a vector. One now sees that Eq. (5.50) is a simultaneous eigenvalue equation for $\mathrm{ad}_{H^{1,2}}$, with $E^{(j)}_\pm$ the eigenvectors corresponding to the eigenvalues $\pm\vec{\alpha}^{\,(j)}$. One has two further eigenvectors, namely $H^1 = I_3$ and $H^2 = \bar{Y}$, both with both eigenvalues equal to zero since $[\vec{H}, H^i] = 0$. The eight eigenvectors $E^{(j)}_\pm$ and $\vec{H}$ form a complete basis for the algebra,[72] and the set $\{\vec{\alpha}^{\,(0)}, \vec{\alpha}^{\,(1)}, \vec{\alpha}^{\,(2)}, \vec{\alpha}^{\,(3)}\}$, with $\vec{\alpha}^{\,(0)} = (0, 0)$, gives all the pairs of simultaneous eigenvalues.

To complete our knowledge of the algebra, i.e., to complete the list of structure constants, we still need the commutators of ladder operators, from which one can obtain the remaining relations satisfied by the t^a, Eq. (5.31). These can of course be computed explicitly, but it is

[72] More precisely, they form a complete basis of the *complexified* algebra, spanned by $\{t^a\}$ if we allow complex coefficients. However, the key point here is completeness, and complexification is immaterial.

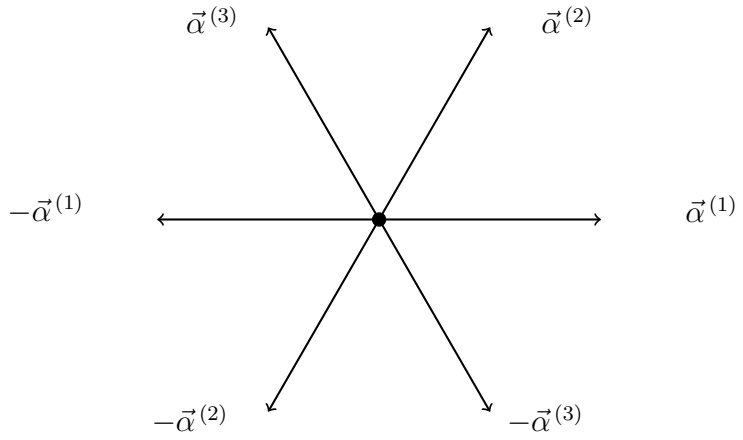

Figure 23: Root system of SU(3).

more instructive to consider the following argument, which also provides a first application of the idea discussed above. Consider the commutator of two ladder operators, and in turn the commutator of this with I_3 and $\bar{Y}$. In the notation of Eq. (5.50) we have, using Jacobi identity Eq. (5.34),

$$\begin{aligned}[\vec{H},[E_s^{(i)},E_t^{(j)}]] &= -[E_t^{(j)},[\vec{H},E_s^{(i)}]] - [E_s^{(i)},[E_t^{(j)},\vec{H}]] \\ &= -s\vec{\alpha}^{\,(i)}[E_t^{(j)},E_s^{(i)}] + t\vec{\alpha}^{\,(j)}[E_s^{(i)},E_t^{(j)}] = (s\vec{\alpha}^{\,(i)} + t\vec{\alpha}^{\,(j)})[E_s^{(i)},E_t^{(j)}]\,,\end{aligned} \tag{5.53}$$

where $s,t=\pm 1$. This means that $[E_s^{(i)},E_t^{(j)}]$ must be a simultaneous eigenvector of ad_{H^1} and ad_{H^2} with eigenvalues given by the components of the vector $(s\vec{\alpha}^{\,(i)} + t\vec{\alpha}^{\,(j)})$. In general, this is *not* a vector of eigenvalues, since only those appearing in Fig. (23) (including the zero vector) are. One then concludes that $[E_s^{(i)},E_t^{(j)}]$ must be zero unless $s\vec{\alpha}^{\,(i)} + t\vec{\alpha}^{\,(j)} = u\vec{\alpha}^{\,(k)}$ for some $u=\pm 1$ and k. In case that such u and k exist, since eigenvalue pairs are non-degenerate, it follows that $[E_s^{(i)},E_t^{(j)}] \propto E_u^{(k)}$. The existence of nontrivial solutions, beside the case $\vec{\alpha}^{\,(i)} - \vec{\alpha}^{\,(i)} = 0$, is guaranteed by the fact that $\vec{\alpha}^{(1)} + \vec{\alpha}^{(3)} = \vec{\alpha}^{(2)}$. Although this argument does not fix the proportionality constant, it will be enough for our purposes.

Diagonal generators As we have already remarked, only two of the Gell-Mann matrices, Eq. (5.36), are diagonal. One can show that for any choice of basis no more than two generators commute and can be diagonalised simultaneously. One then says that the group SU(3) is of *rank 2*. From the physical point of view, if SU(3) were an exact symmetry of nature, then only two generators could be measured simultaneously, but there would be a large arbitrariness in choosing them (e.g., any linear combination of I_1 and $\bar{Y}$ would do). This is analogous to the possibility of measuring only one angular momentum component at a time – but any component is allowed. In practice, however, a preferred choice exists for which generators of SU(3) to use to label the physical states: indeed, we know that strong interactions break SU(3) down to $\mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$, so we choose I_3 and Y. On top of these, $\vec{I}^2$ can be also used to label particles, since it commutes with I_3 and, obviously, with Y.[73]

[73]Recall that the "total isospin" $\vec{I}^2$ is *not* an element of the Lie algebra of the group, see footnote 56.

5.2.2 Representations of SU(3)

We discuss now the representation of SU(3), first from a general perspective, and then focussing on the representations of lowest dimension, looking for multiplets matching the baryon and meson multiplets observed in nature.

As in the case of SU(2), it is easier to build representations of SU(3) by exponentiating representations of its algebra. Since we are interested in unitary representations of the group, we look for Hermitian representations of the algebra.[74] These are obtained if we can find a set of eight Hermitian matrices obeying the commutation relations Eq. (5.31), or equivalently a set of matrices representing the ladder operators, obeying $I_\pm^\dagger = I_\mp$ and similarly for the others, and two Hermitian matrices representing I_3, and Y, satisfying the commutation relations Eq. (5.47), (5.48), as well as the remaining relations involving pairs of ladder operators. Since I_3 and Y are Hermitian and commute, we can always choose the basis of our representation so that they are diagonal. In other words, we can take as a basis of our representation space a complete set of simultaneous eigenvectors of the representatives of I_3 and Y. Representations are then characterised by the corresponding pairs of eigenvalues, called *weights* or *weight vectors* (which in general *do not* coincide with the root vectors), and can be represented graphically as diagrams in the (I_3, Y) plane (mathematicians would rather use the $(I_3, \bar{Y}) = (I_3, \frac{\sqrt{3}}{2}Y)$ plane), known as *weight diagrams*. These completely and uniquely characterise the representation, if information is included concerning the degeneracy of each weight. For practical purposes we would also like to know how the eigenvectors are organised in isospin multiplets, i.e., multiplets with a constant value of $\vec{I}^2$.

As already pointed out above, the choice to diagonalise I_3 and Y is motivated from the fact that they correspond to the quantum numbers used to classify hadrons. Finding diagrams that match the baryon octet or any other multiplet of hadrons would indicate that we are on the right track and we have correctly identified the symmetry. We have already remarked that if SU(3) were truly a symmetry, any basis would be equivalent, and no choice of states would be "more physical" than any other. The reason why I_3 and Y, and not any other pair of generators, is the most convenient, and physical adequate, choice is that SU(3) is actually broken by sizeable strong-interaction effects down to $\mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$, and furthermore that also the full $\mathrm{SU}(2)_I$ is broken by electromagnetic interactions, as well as by small strong-interaction effects, down to $U(1)_{I_3}$. This singles out the basis of physical particles with definite strangeness and electric charge, or equivalently hypercharge and third isospin component, as well as baryon number, since only $U(1)_B \times U(1)_S \times U(1)_{\mathrm{em}} = U(1)_B \times U(1)_Y \times U(1)_{I_3}$ is a true symmetry of Nature (as long as weak interactions are ignored).

5.2.3 The simplest representations of SU(3)

While we could work out rather easily all the irreducible representations of SU(2), the case of SU(3) is more intricate. In general, one starts from an eigenvector annihilated by the three raising operators I_+, V_+, W_+ (which must exist in a finite dimensional representation), and build the rest using the descent operators I_-, V_-, W_-. This requires care since the same eigenvalue pair can be reached following different routes, and one must make sure that all and only the

[74] Since SU(3) is simply connected, the representations obtained by exponentiating its algebra are regular representations, not projective. Moreover, since SU(3) is compact, all its finite-dimensional representations are equivalent to unitary representations.

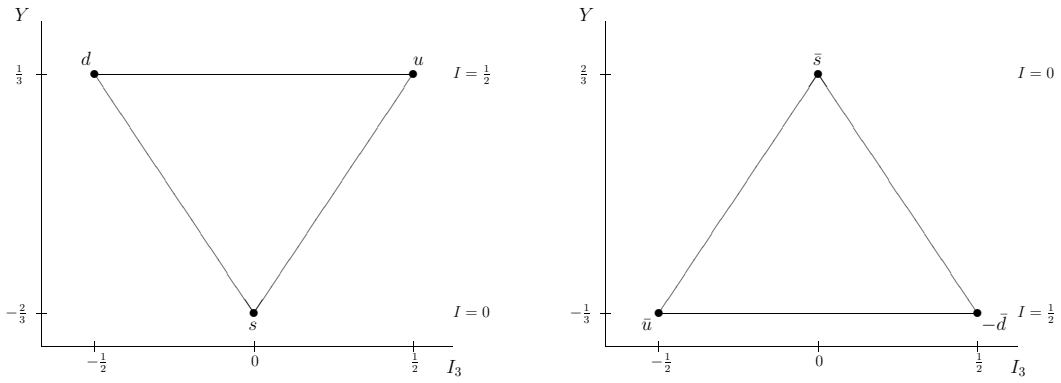

Figure 24: Weight diagrams of the fundamental (**3**) and complex conjugate ($\bar{\mathbf{3}}$) representations of SU(3).

linearly independent possibilities are taken into account. Instead of doing this, we will first follow an easier (and in a sense reversed) procedure by looking at the easy-to-guess representations of the group first, and obtaining the weight diagram afterwards.

Trivial representation For any group, the simplest representation is $D_T(g) = 1$ for all g: this is the *trivial representation*, which is one-dimensional, and therefore denoted by **1**, and also called the *singlet* representation. Correspondingly, all the elements of the Lie algebra of the group are represented by zero. The weight diagram contains a single point at $(I_3, Y) = (0, 0)$, which obviously corresponds to an isospin singlet.

Fundamental representation For any matrix Lie group like SU(3), the second simplest representation is that provided by the matrix group itself. This is the *fundamental* (or more precisely *defining*) representation, $D_F(U) = U$, to which corresponds the fundamental representation of the algebra with generators $T_a^{(F)}$. Formally,

$$D_F(U(\boldsymbol{\alpha})) = U(\boldsymbol{\alpha}) = e^{i\boldsymbol{\alpha}\cdot\boldsymbol{t}} \Longrightarrow T_a^{(F)} = t^a \,. \tag{5.54}$$

By construction, $T_a^{(F)}$ obey the commutation relations Eq. (5.31) (as well as the normalisation condition Eq. (5.32)). This representation is three-dimensional, so denoted by **3** and also called the *triplet* representation. The algebra is conveniently represented in terms of the generators t^a obtained from the Gell-Mann matrices Eq. (5.36), which display diagonal I_3 and Y. It is easy to read off the weights: these are

$$(I_3, Y) = (\tfrac{1}{2}, \tfrac{1}{3}), (-\tfrac{1}{2}, \tfrac{1}{3}); (0, -\tfrac{2}{3}) \,, \tag{5.55}$$

corresponding to the canonical basis vectors $e^{(j)}$, $e_i^{(j)} = \delta_{ij}$, $j = 1, 2, 3$. It is easy to see, looking at the form of $I_{1,2} = t^{1,2}$, that $e^{(1,2)}$ form an isodoublet, while $e^{(3)}$ is an isosinglet. The weight diagram of the **3** representation is shown in Fig. 24.

Complex-conjugate representation Another representation that is easy to spot is the *complex-conjugate* or *antifundamental* representation,

$$D_C(U(\boldsymbol{\alpha})) = U(\boldsymbol{\alpha})^* = e^{i\boldsymbol{\alpha}\cdot(-\boldsymbol{t}^*)} \Longrightarrow T_a^{(C)} = (-t^a)^* \,. \tag{5.56}$$

This is clearly a valid representation since $D_C(U_1U_2) = (U_1U_2)^* = U_1^*U_2^* = D_C(U_1)D_C(U_2)$. This implies that the matrices $T_a^{(C)} = -(t^a)^* = -(t^a)^T$ obey Eq. (5.31), as one can easily see explicitly: taking the complex conjugate of this equation, we find

$$[T_a^{(C)}, T_b^{(C)}] = [(-t^a)^*, (-t^b)^*] = [t^a, t^b]^* = -if_{abc}(t^c)^* = if_{abc}(-t^c)^* = if_{abc}T_c^{(C)}\,. \tag{5.57}$$

These generators still obey the normalisation condition Eq. (5.32) since clearly $\mathrm{tr}\, T_a^{(C)}T_b^{(C)} = [\mathrm{tr}\, T_a^{(F)}T_b^{(F)}]^* = \frac{1}{2}\delta^{ab}$. Also in this case the representatives of I_3 and Y are diagonal, so that the canonical basis vectors $e^{(j)}$ are their eigenvectors, but since $T_{3,8}^{(C)} = -t^{3,8}$ this time the corresponding weights are

$$(I_3, Y) = (-\tfrac{1}{2}, -\tfrac{1}{3}), (\tfrac{1}{2}, -\tfrac{1}{3}); (0, \tfrac{2}{3})\,. \tag{5.58}$$

This shows that, differently from the case of SU(2), for SU(3) the fundamental and the complex conjugate representations are not equivalent. For this reason one uses the notation $\bar{\mathbf{3}}$ for the complex-conjugate representation. Finally, as in the fundamental representation, $e^{(1,2)}$ form an isodoublet (although their eigenvalues have changed sign) and $e^{(3)}$ is an isosinglet. The weight diagram of the $\bar{\mathbf{3}}$ representation is shown in Fig. 24. In general, the complex conjugate $\bar{\boldsymbol{R}}$ of a representation $\boldsymbol{R}$, obtained by taking the complex conjugate of the representative matrices, $D_{\bar{\boldsymbol{R}}}(U) = D_{\boldsymbol{R}}(U)^*$, has a weight diagram obtained by reflecting that of $\boldsymbol{R}$ both horizontally and vertically.

Adjoint representation As we saw already for SU(2), a Lie group can be represented on its own algebra (seen as a linear space) via the adjoint representation (see Section 4.2.4, p. 94). Recall that the algebra of SU(3) is the vector space of Hermitian 3×3 traceless matrices A equipped with the commutator. For any given $U \in \mathrm{SU}(3)$ consider the mapping

$$\mathrm{Ad}_U A \equiv UAU^\dagger\,. \tag{5.59}$$

For any U, Ad_U is a linear operator from the algebra into itself, and moreover

$$\mathrm{Ad}_{U_2}\mathrm{Ad}_{U_1}A = \mathrm{Ad}_{U_2}U_1AU_1^\dagger = U_2U_1AU_1^\dagger U_2^\dagger = \mathrm{Ad}_{U_2U_1}A\,, \tag{5.60}$$

i.e., Ad_U provides a representation of the group, using the group algebra as the representation space. To find the associated representation of the algebra, one considers infinitesimal transformations $U \simeq 1 + i\epsilon_a t^a$, and using the definition Eq. (5.59) one finds

$$\mathrm{Ad}_U A \simeq A + i\epsilon_a[t^a, A] = A + i\epsilon_a\,\mathrm{ad}_{t^a}A\,. \tag{5.61}$$

where ad_X was defined in Eq. (5.52). Exactly as for SU(2) [see Eqs. (4.111) and (4.112)] the operators ad_{t^a} automatically satisfy the commutation relations Eq. (5.31) as consequence of the Jacobi identity Eq. (5.34),

$$\begin{aligned}[\mathrm{ad}_{t^a}, \mathrm{ad}_{t^b}]A &= [t^a, [t^b, A]] - [t^b, [t^a, A]] = [[t^a, t^b], A] + [t^b, [t^a, A]] - [t^b, [t^a, A]] \\ &= [[t^a, t^b], A] = \mathrm{ad}_{[t^a, t^b]}A = if_{abc}[t^c, A] = if_{abc}\mathrm{ad}_{t^c}A\,,\end{aligned} \tag{5.62}$$

holding for any A in the algebra, and so

$$[\mathrm{ad}_{t^a}, \mathrm{ad}_{t^b}] = \mathrm{ad}_{[t^a, t^b]} = if_{abc}\mathrm{ad}_{t^c}A\,. \tag{5.63}$$

It may perhaps be easier to work in components to see that the Ad_U and ad_{t^a} do provide us with representative matrices for the elements of the group and of the algebra, respectively. Writing an element A of the algebra in components, $A = A_a t^a$, with $A_a = 2\,\mathrm{tr}\, t^a A$, one finds

$$(\mathrm{Ad}_U A)_a = 2\,\mathrm{tr}\, t^a U A U^\dagger = \left(2\,\mathrm{tr}\, t^a U t^b U^\dagger\right) A_b \equiv \mathcal{U}(U)_{ab} A_b\,, \tag{5.64}$$

where the matrices $\mathcal{U}$ satisfy

$$\mathcal{U}(U_2U_1)_{ab} A_b = (\mathrm{Ad}_{U_2U_1} A)_a = (\mathrm{Ad}_{U_2}\mathrm{Ad}_{U_1} A)_a = \mathcal{U}(U_2)_{ab}\,\mathcal{U}(U_1)_{bc} A_c\,, \tag{5.65}$$

i.e., they give a representation of the group. In particular, $\mathcal{U}(U^\dagger) = \mathcal{U}(U^{-1}) = \mathcal{U}(U)^{-1}$. Using the explicit expression Eq. (5.64), we find that

$$\begin{aligned} (\mathcal{U}(U)_{ab})^* &= 2\,\mathrm{tr}\, U t^b U^\dagger t^a = 2\,\mathrm{tr}\, t^a U t^b U^\dagger = \mathcal{U}(U)_{ab}\,, \\ \mathcal{U}(U)^T_{ab} &= \mathcal{U}(U)_{ba} = 2\,\mathrm{tr}\, t^b U t^a U^\dagger = 2\,\mathrm{tr}\, t^a U^\dagger t^b U = \mathcal{U}(U^\dagger)_{ab} = \mathcal{U}(U)^{-1}_{ab}\,, \end{aligned} \tag{5.66}$$

so that the representation is orthogonal, i.e., unitary and real. For the representation of the algebra, we find in components

$$\begin{aligned} (\mathrm{ad}_{t^a} A)_b &= 2\,\mathrm{tr}\, t^b [t^a, A] = 2\,\mathrm{tr}\,(t^b [t^a, t^c]) A_c = 2\,\mathrm{tr}\,(t^b t^d) i f_{acd} A_c = i f_{acb} A_c \\ &= -i f_{abc} A_c \equiv (T^{(A)}_a)_{bc} A_c\,, \end{aligned} \tag{5.67}$$

with the Hermitian matrices $T^{(A)}_a$ (since antisymmetric and purely imaginary) providing the adjoint representation of the algebra. To see explicitly that they obey the commutation relations Eq. (5.31), one exploits the properties of the structure constants to recast Eq. (5.35), which follows from the Jacobi identity, as

$$\begin{aligned} (-i f_{bam})(-i f_{cmn}) - (-i f_{cam})(-i f_{bmn}) &= i f_{bcm}(-i f_{man})\,, \\ (T^{(A)}_b)_{am}(T^{(A)}_c)_{mn} - (T^{(A)}_c)_{am}(T^{(A)}_b)_{mn} &= i f_{bcm}(T^{(A)}_m)_{an}\,, \\ ([T^{(A)}_b, T^{(A)}_c])_{an} &= i f_{bcm}(T^{(A)}_m)_{an}\,. \end{aligned} \tag{5.68}$$

The normalisation of the matrices $T^{(A)}_a$ follows from Eq. (5.37):

$$\mathrm{tr}\, T^{(A)}_a T^{(A)}_b = -f_{amn} f_{bnm} = f_{amn} f_{bmn} = 3\delta_{ab}\,. \tag{5.69}$$

Our results concerning the adjoint representation are summarised as follows:

$$D_A(U(\boldsymbol{\alpha})) = \mathrm{Ad}_{U(\boldsymbol{\alpha})} = e^{i\mathrm{ad}_{\boldsymbol{\alpha}\cdot\boldsymbol{t}}} \Longrightarrow T^{(A)}_a = \mathrm{ad}_{t^a}\,. \tag{5.70}$$

The adjoint representation is eight-dimensional, since there are eight generators t^a, and is denoted by **8**. Since it is a real representation it equals its complex conjugate, so $\overline{\mathbf{8}} = \mathbf{8}$.

We now get to the point, which is the weight diagram of this representation. Recalling that the weight diagram represents the pairs of simultaneous eigenvalues of the representatives of I_3 and Y, it is enough to look back at Eqs. (5.47) and (5.50): the simultaneous eigenvectors of $I^{(A)}_3 = \mathrm{ad}_{t^3}$ and $Y^{(A)} = \mathrm{ad}_{\frac{2}{\sqrt{3}}t^8}$ are just the ladder operators, whose eigenvalues can be read off Eq. (5.47) (or suitably rescaling the root vectors, Eq. (5.51)), together with the diagonal

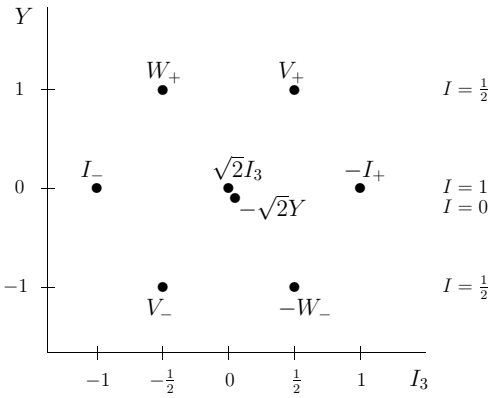

Figure 25: Adjoint representation of SU(3).

generators which commute with each other: The weights and the corresponding eigenvectors are then

$$\begin{array}{rccccc} (I_3, Y) = & (\frac{1}{2}, 1), (-\frac{1}{2}, 1); & (1,0), (0,0), (-1,0); & (0,0); & (\frac{1}{2}, -1), (-\frac{1}{2}, -1)\,, \\ & V_+, W_+; & I_+, I_3, I_-; & Y; & W_-, V_-\,. \end{array} \tag{5.71}$$

These form respectively an isodoublet, an isotriplet, an isosinglet, and another isodoublet. The weight diagram of the adjoint representation is shown in Fig. 25, and a quick look shows that it corresponds precisely to the baryon octet, Fig. 22 (top right). We have then shown that there are irreducible representations of SU(3) that correspond to hadronic multiplets, and so SU(3) can indeed be the bigger symmetry group we were looking for.

5.2.4 More general representations of SU(3): the "eightfold way"

To finally be satisfied with SU(3), the next task is to find a representation that accommodates the baryonic resonances, see Fig. 22 (bottom). To do this, one has to go beyond the simple representations discussed above and look a bit into how general representations are constructed. The idea is similar to that employed for SU(2): since the ladder operators make new eigenvectors out of the existing ones, if the representation has to be finite-dimensional then at some point one has to stop, and one will find an eigenstate of I_3 and Y that is annihilated by the three raising operators I_+, V_+, W_+. The other eigenvectors will then be obtained from this one by applying the lowering operators I_-, V_-, W_-. Since the same pair of eigenvalues can be obtained in different ways, one has to make sure to count all the independent eigenvectors with the same weight. The detailed implementation of this procedure is clearly more complicated than for SU(2), but we will see that it can be worked out rather easily in the case we are interested in.

Decuplet representation Suppose that an eigenvector $|D\rangle$ exists with the isospin quantum numbers of the Δ^{++}, i.e., $I = I_3 = \frac{3}{2}$, which is furthermore an eigenvector of Y. If it is part of a representation that looks like Fig. 22 (bottom), then it must be annihilated by I_+, V_+, W_+, as well as by W_-. Being an eigenvector of I_3 and Y it is also an eigenvector of V_3 and of W_3; in particular, it is an $\mathrm{SU}(2)_W$ singlet, $W = W_3 = -\frac{1}{2}I_3 + \frac{3}{4}Y = 0$. This tells us that $Y = 1$ for $|D\rangle$. Since $I = \frac{3}{2}$, applying I_- repeatedly we obtain three more states with $I_3 = \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$, the last of which is annihilated by I_-. All these states have the same hypercharge 1 as $|D\rangle$ since

$[Y, I_-] = 0$, so they are also W_3 and Y_W eigenstates with W_3 eigenvalues $\frac{3}{4} - \frac{1}{4} = \frac{1}{2}$, $\frac{3}{4} + \frac{1}{4} = 1$, and $\frac{3}{4} + \frac{3}{4} = \frac{3}{2}$, and Y_W eigenvalues $-\frac{3}{2} - \frac{1}{2} = -2$, $-\frac{3}{2} + \frac{1}{2} = -1$, and $-\frac{3}{2} + \frac{3}{2} = 0$. Since $[W_+, I_-] = 0$ (there is no root vector equal to $-\alpha^{(1)} + \alpha^{(3)}$), all these states are also annihilated by W_+; since $[V_+, I_-] \propto W_+$ (since $-\alpha^{(1)} + \alpha^{(2)} = \alpha^{(3)}$), these are also annihilated by V_+, and so there are no states with hypercharge larger than 1.

A representation of SU(3) must also provide a representation of any of its subgroups, in particular $\mathrm{SU}(2)_{I,V,W}$. An $\mathrm{SU}(2)_I$ rotation of π, $e^{i\pi I_1}$, applied on an eigenvector of I_3 and Y, gives a new eigenvector with opposite eigenvalue of I_3 and the same eigenvalue of Y. This means that the weight diagram of a representation should be symmetric under a reflection through the axis $I_3 = 0$. The same applies to an $\mathrm{SU}(2)_W$ rotation of π, $e^{i\pi W_1}$, that sends an eigenvector of $W_3 = -\frac{1}{2}I_3 + \frac{3}{4}Y$ and $Y_W = -I_3 - \frac{1}{2}Y$ to another eigenvector with opposite eigenvalue of W_3 and the same eigenvalue of Y_W. The weight diagram must then be symmetric under a reflection through the axis $W_3 = 0$, i.e., $I_3 = \frac{3}{2}Y$. There must therefore be three more states with W_3 eigenvalues $-\frac{1}{2}$, -1, and $-\frac{3}{2}$, and Y_W eigenvalues -2, -1, and 0, which translate into eigenvalues of I_3 and Y equal to $(I_3, Y) = (1, 0); (\frac{1}{2}, -1); (0, -2)$. Using now the symmetry under reflection through the axis $I_3 = 0$, we find two more states with $(I_3, Y) = (-1, 0); (-\frac{1}{2}, -1)$, which completes the outline of an equilateral triangle of side length 3. Notice that this is symmetric also under a reflection through the axis $V_3 = \frac{1}{2}I_3 + \frac{3}{4}Y = 0$, as it should be since the weight diagram must be symmetric under an $\mathrm{SU}(2)_V$ rotation of π.

Since the four weights $(I_3, Y) = (\frac{3}{2}, 1); (\frac{1}{2}, 1); (-\frac{1}{2}, 1); (-\frac{3}{2}, 1)$, from which we started, are non-degenerate, all the weights on the perimeter of the triangle are non-degenerate. Using repeatedly the descent operators I_-, V_-, W_- on $|D\rangle$, one can build in several ways an eigenvector with weight $(I_3, Y) = (0, 0)$: $I_-V_-|D\rangle$, $V_-I_-|D\rangle$, and $W_-I_-^2|D\rangle$. But $[I_-, V_-] = 0$ (there is no root vector equal to $-\alpha^{(1)} - \alpha^{(2)}$), so the first two are the same vector, and $[I_-, W_-] = -cV_-$ for some constant c (since $-\alpha^{(1)} - \alpha^{(3)} = -\alpha^{(2)}$), and so

$$W_-I_-^2|D\rangle = (I_-W_- + cV_-)I_-|D\rangle = cV_-I_-|D\rangle + I_-(I_-W_- + cV_-)|D\rangle = 2cV_-I_-|D\rangle \qquad (5.72)$$

is proportional to the same vector. The weight $(I_3, Y) = (0, 0)$ is then non-degenerate as well. Since the eigenvectors with weights $(I_3, Y) = (\frac{3}{2}, \frac{3}{2}); (1, 1); (\frac{1}{2}, \frac{1}{2}); (0, 0)$ are annihilated by I_+ (there are no weights on their right), they are also eigenvectors of $\vec{I}^2$ with $I = \frac{3}{2}, 1, \frac{1}{2}, 0$. These vectors can be reached from $|D\rangle$ by repeatedly applying V_-; in fact, since $|D\rangle$ is an eigenvector of V_3 and of $\vec{V}^2$ with eigenvalues $V_3 = \frac{1}{2}I_3 + \frac{3}{4}Y = \frac{3}{2}$ and $V(V+1)$, with $V = V_3\frac{3}{2}$, they form a quartet of $\mathrm{SU}(2)_V$. The eigenvectors on their left can be reached by repeatedly applying I_-, and so they complete the corresponding isospin multiplets.

Summarising, we have found an irreducible representation containing:

- four states $I_-^{0,1,2,3}|D\rangle$ forming an isoquartet, $I = \frac{3}{2}$, with $Y = 1$;
- three states $I_-^{0,1,2}V_-|D\rangle$ forming an isotriplet, $I = 1$, with $Y = 0$;
- two states $I_-^{0,1}V_-^2|D\rangle$ forming an isodoublet, $I = \frac{1}{2}$, with $Y = -1$;
- one state $V_-^3|D\rangle$ forming an isosinglet, $I = 0$, with $Y = -2$.

This is a total of ten states, forming the *decuplet*, or **10**, representation. These perfectly reproduced the quantum numbers of the known baryonic resonances, except for the last state which did not have an experimental counterpart.

The "eightfold way" If SU(3) symmetry really is the explanation behind the observed hadronic multiplets, then all the hadronic multiplets should fit into irreducible representations, not just the lightest baryons. We have found an 8-dimensional representation (the adjoint representation), but one can show that there are no 7-dimensional and 9-dimensional irreducible representations – we only found a 10-dimensional one. At this point, one is faced with a choice: either discard SU(3) as the correct symmetry, or make a bold move and predict the existence of new hadrons that would fill the vacant places in the available representations. The new particles needed to complete the multiplets were another light meson for the meson octet, and another baryonic resonance for the baryon decuplet.

The missing meson needed to be a $J^P = 0^-$ pseudoscalar meson, and have quantum numbers $I = Y = 0$, and so an expected electric charge $Q = 0$ from the Gell-Mann–Nishijima formula. This was found experimentally in 1961 [12], and is known as the η meson. The missing baryon needed to have spin $\frac{3}{2}$, isospin $I = 0$ and strangeness $S = Y - B = -3$, and so $Q = -1$ from the Gell-Mann–Nishijima formula. A look at the mass pattern of the resonances, $m_\Delta = 1232$ MeV, $m_{\Sigma^*} = 1384$ MeV, $m_{\Xi^*} = 1533$ MeV, and the corresponding mass splittings $m_{\Sigma^*} - m_\Delta = 152$ MeV and $m_{\Xi^*} - m_{\Sigma^*} = 149$ MeV, leads to the empirical relation $m(S) = m(\Delta) + 150\text{ MeV} \cdot |S|$, and so to an educated guess of $m \sim 1682$ MeV for the missing baryon.[75] Notice that a similar mass formula approximately works for the baryon octet as well. In 1964 a spin-$\frac{3}{2}$ baryon resonance with the predicted properties, named Ω^-, was indeed observed at a mass $m_\Omega = 1672$ MeV [13].

With the η and the Ω^- completing the respective multiplets, the classification of hadron multiplets in terms of irreducible representations of SU(3) was successfully achieved. This classification had been proposed by Murray Gell-Mann in 1961, under the rather bizarre name of *eightfold way*, and independently by Yuval Ne'eman, also in 1961 [14]. An important role in suggesting SU(3) as the relevant approximate symmetry of the strong interactions was played by the model of composite hadrons of S. Sakata and its developments.

5.2.5 From SU(3) invariance to the quark model

Although the "eightfold way" allowed to nicely classify the existing hadrons, and even predict correctly the existence of new ones, it had nonetheless a number of issues. The most evident shortcoming was that, while from the group-theoretical point of view all representations are good, Nature showed only a small subset of them. For baryons, only octets (like the lightest spin-$\frac{1}{2}$ baryons) and decuplets (like the spin-$\frac{3}{2}$ baryonic resonances) were observed; for mesons, only octets (like the lightest pseudoscalar mesons K, π and η) and singlets (like the pseudoscalar meson η'). Representation theory alone could not explain why only certain irreducible representations appeared in nature, while other did not.

Quite surprisingly, the fundamental and complex conjugate representations (the next-to-simplest representations) did not appear among the hadronic multiplets. On the other hand, it is a well-known result in the theory of representations of SU(3) that all irreducible representations can be obtained by reducing tensor products of fundamental ($\mathbf{3}$) and complex conjugate ($\overline{\mathbf{3}}$) representations.[76] The easiest way to work out the composition of SU(3) representations is to do it graphically by using the weight diagrams, taking into account that I_3 and Y are additive

[75] This estimate was done actually through a more accurate formula, the Gell-Mann–Okubo formula, discussed below in Section 5.2.7.

[76] Actually the fundamental representation suffices, as we will see below.

quantities, and so the tensor product of two states $|i_3^{(1)}, y^{(1)}\rangle$ and $|i_3^{(2)}, y^{(2)}\rangle$ will have eigenvalues of I_3 and Y equal to the sum of the eigenvalues of those two states, i.e., $i_3^{(1)} + i_3^{(2)}$ and $y^{(1)} + y^{(2)}$.

Take two representations R_1 and R_2 with generators represented by Hermitian matrices $T_a^{(R_{1,2})}$, and such that $T_3^{(R_{1,2})}$ and $T_8^{(R_{1,2})}$ are diagonal. The representation spaces are spanned by the simultaneous eigenvectors $|i_3^{(1)}, y^{(1)}\rangle$ and $|i_3^{(2)}, y^{(2)}\rangle$ of I_3 and Y, i.e., of $T_3^{(R_{1,2})}$ and $T_8^{(R_{1,2})}$. The tensor product representation has representation space spanned by $|i_3^{(1)}, y^{(1)}; i_3^{(2)}, y^{(2)}\rangle = |i_3^{(1)}, y^{(1)}\rangle \otimes |i_3^{(2)}, y^{(2)}\rangle$, and generators represented by $T_a^{(R_1 \otimes R_2)} = T_a^{(R_1)} \otimes \mathbf{1}^{(R_2)} + \mathbf{1}^{(R_1)} \otimes T_a^{(R_2)}$. Clearly,

$$\begin{aligned} T_3^{(R_1\otimes R_2)}|i_3^{(1)}, y^{(1)}; i_3^{(2)}, y^{(2)}\rangle &= (T_3^{(R_1)} \otimes \mathbf{1}^{(R_2)} + \mathbf{1}^{(R_1)} \otimes T_3^{(R_2)})|i_3^{(1)}, y^{(1)}\rangle \otimes |i_3^{(2)}, y^{(2)}\rangle \\ &= (T_3^{(R_1)}|i_3^{(1)}, y^{(1)}\rangle) \otimes (\mathbf{1}^{(R_2)}|i_3^{(2)}, y^{(2)}\rangle) \\ &\quad + (\mathbf{1}^{(R_1)}|i_3^{(1)}, y^{(1)}\rangle) \otimes (T_3^{(R_2)}|i_3^{(2)}, y^{(2)}\rangle) = (i_3^{(1)} + i_3^{(2)})|i_3^{(1)}, y^{(1)}; i_3^{(2)}, y^{(2)}\rangle\,, \end{aligned}$$

and similarly for $T_8^{(R_1\otimes R_2)}$.

This is depicted in Fig. 26 for the tensor product of a fundamental and an antifundamental representation: we just draw a triangle corresponding to the first $\mathbf{3}$ in the (I_3, Y)-plane, and three more triangles corresponding to a $\bar{\mathbf{3}}$ representation but centred on the vertices of the first one. The coordinates of the vertices of the "outer" triangles correspond to the wieghts, i.e., the eigenvalue pairs of I_3 and Y, appearing in the $\bar{\mathbf{3}} \otimes \mathbf{3}$ representation, obtained by adding a weight of the fundamental representation and a weight of the antifundamental representation. We then decompose the result into irreducible representations: here the task is easy as one easily identifies an octet, leaving a single state that can only be a singlet. One then concludes that $\bar{\mathbf{3}} \otimes \mathbf{3} = \mathbf{8} \oplus \mathbf{1}$. Remarkably, these are precisely the representations appearing in mesonic multiplets.

One proceeds similarly for the product of two fundamental representations, drawing now a triangle corresponding to the first $\mathbf{3}$ representation and three more triangles centred on the vertices of the first one corresponding to the second $\mathbf{3}$ representation (see Fig. 27). In this case, in decomposing the result into irreducible representations we are helped by the fact that there is an easily recognisable set of states corresponding to an antifundamental representation $\bar{\mathbf{3}}$, and once that is removed one is left with six non-degenerate states. An argument similar to the one used to build the decuplet shows that this is one of the irreducible representations of SU(3), the $\mathbf{6}$. In assigning states in degenerate (I_3, Y)-eigenspaces to different irreducible representations one can also take advantage of symmetry considerations: in the case at hand, the pair of eigenvalues corresponding to the weight in the top right corner of the diagram comes from the symmetric state $|\frac{1}{2}, \frac{1}{3}\rangle \otimes |\frac{1}{2}, \frac{1}{3}\rangle$, and since the descent operators do not change the symmetry properties under exchange of the two factors in the product states, one will obtain an irreducible representation of symmetric states – the two-index symmetric representation, or $\mathbf{6}$. For the degenerate weights there is a second combination of product states that is antisymmetric under exchange, which forms a second irreducible representation – the two-index antisymmetric representation, which coincides with the $\bar{\mathbf{3}}$. In conclusion, we found $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \bar{\mathbf{3}}$. A similar procedure shows that $\mathbf{6} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8}$, where again one uses symmetry properties to assign states corresponding to degenerate weight to one irreducible representation or the other. If we now consider the tensor product of three fundamental representations, using the results collected so far we find

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = (\mathbf{6} \oplus \bar{\mathbf{3}}) \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}\,. \tag{5.73}$$

Interestingly, we obtain the representations appearing in baryonic multiplets, except for the presence of a singlet representation that does not occur in nature.

A "natural" interpretation of the results discussed above is that hadrons are not elementary, but rather composite particles, built out of constituents that transform in the fundamental or antifundamental representation of SU(3): three fundamental constituents for a baryon (and three antifundamental for an antibaryon), and a fundamental and an antifundamental constituent for a meson.[77] This interpretation had a number of upsides. First of all, it immediately reduces the problem of the large zoo of supposedly elementary hadrons to a much more manageable number of beasts. Secondly, it allows to explain the approximate mass relation $m(S) = m(0) + 150\ \text{MeV} \cdot |S|$ valid within the octet and decuplet baryon multiplets in terms of the masses of the constituents. Finally, the Gell-Mann–Nishijima formula $Q = I_3 + \frac{1}{2}Y = I_3 + \frac{1}{2}(B + S)$ for the electric charge of hadrons, which is a relation between additive quantities, will automatically hold for any hadron if it holds for the constituents.

This interpretation also had a considerable number of downsides. As we saw above, it *almost* solves the riddle of the SU(3) representations appearing in nature, but the price to pay is to introduce objects that transform according to representations that do *not* appear in nature. As soon as the new constituents are introduced, one has to postulate that they have to stay hidden away forever inside hadrons – a very suspicious move. Nonetheless, if one assumes that quarks are the elementary constituents of hadrons, one can hope to fully solve the representation puzzle using the constraints coming from the exchange symmetry of identical particles, which might forbid the unobserved representations. However, as we will see below, this actually leads to serious theoretical problems. As it turns out, these problems can be solved (essentially by introducing again something that cannot be observed...), and their solution led to the modern theory of strong interactions, i.e., QCD. The elementary constituents became known as *quarks*, another bizarre name courtesy of M. Gell-Mann, who at first used them only as fictitious particles to explain the representations found in hadronic multiplets. The first one to believe in the physical existence of the hadronic constituents was George Zweig, who called them "aces" instead, and who developed a nice and correct but quickly dismissed model of hadrons based on them.

Quark quantum numbers If the idea of quarks and antiquarks as elementary constituents of hadrons is viable, one should be able to assign them appropriate values of the various quantum numbers, that reproduce the observed phenomenology. By construction, quarks correspond to the simultaneous eigenstates of I_3 and Y in the fundamental representation.[78] There are then three types or *flavours* of quarks: these are denoted as u ("up"), d ("down") and s ("strange"), with isospin and hypercharge assigned according to Fig. 24, or Table 11. Quantum Field Theory requires the existence of an antiparticle for each particle, so we have to introduce the *antiquarks* $\bar{u}$, $\bar{d}$, and $\bar{s}$, to which we must assign the same spin and mass as the quarks, and minus all the

[77] This all looks natural from a modern perspective, but the first proposals in this direction either did not work out properly, as in Fermi and Yang's approach, or were initially not accepted, as was the case with Zweig's proposal which turned out to be correct after all.

[78] As already pointed out, if SU(3) were an exact symmetry, then any linear superposition of these eigenvectors would look the same to strong interactions. SU(3) is instead broken down to (approximate) $\text{SU}(2)_I \times \text{U}(1)_Y$, so that only isospin and hypercharge eigenstates are "good" states, whose superposition is forbidden by superselection rules.

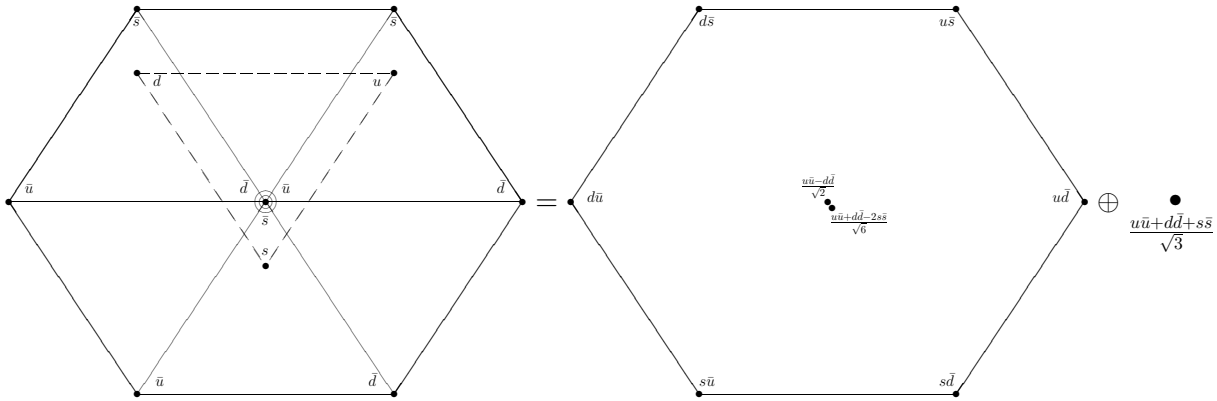

Figure 26: Composing representations of SU(3): $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$.

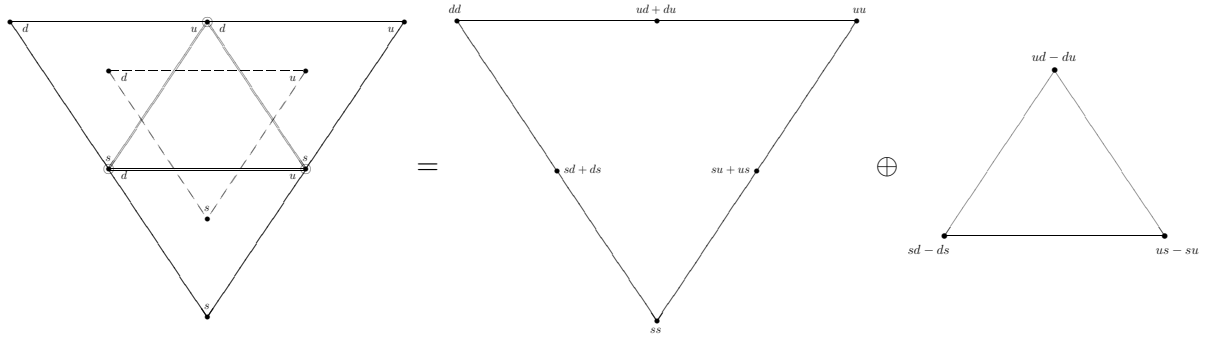

Figure 27: Composing representations of SU(3): $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \bar{\mathbf{3}}$.

charges, such as I_3 and Y.[79]

Irrespectively of the precise form of the linear combinations of products of three quark states (or of a quark and an antiquark state), corresponding to a member of an irreducible representation, the quark content of each hadron automatically follows from the corresponding values of I_3 and Y. The values of these quantum numbers for the composite states are then fixed entirely by the representation, and can be read off Figs. 26 and 27. For example, the proton is a quark-quark-quark state in the top right corner of the octet. In Fig. 26 this corresponds to a $u\bar{s}$ state in the $\mathbf{3} \otimes \bar{\mathbf{3}}$ product;[80] from Fig. 27 we see that the $\bar{s}$ state has the same quantum numbers of a particular combination of ud and du states, so that the proton has quark content uud. The quark content of octet and decuplet baryons is shown in Fig. 28, while that of octet and singlet mesons is shown in Fig. 29.

Conversely, one can extract the quark composition of a hadron from its quantum numbers I_3

[79] More precisely, QFT demands that if a particle transforms in a certain representation of some symmetry group, then its antiparticle must transform in the corresponding complex-conjugate representation, see footnote 65. One can always associate particle and antiparticle with the same representative vector $e^{(i)}$, so that for each flavour charge conjugation is implemented as $Cf = \bar{f}$ – we could introduce a phase (or worse) between the representative vectors for particle and antiparticle, which would result in a corresponding phase in the C transformation, but why make our life more complicated, when we are free to choose the charge-conjugation phase for non-self-conjugate particles?

[80] Notice that this is meant only for representation theoretic purposes, and does not mean that the proton is quark-antiquark states.

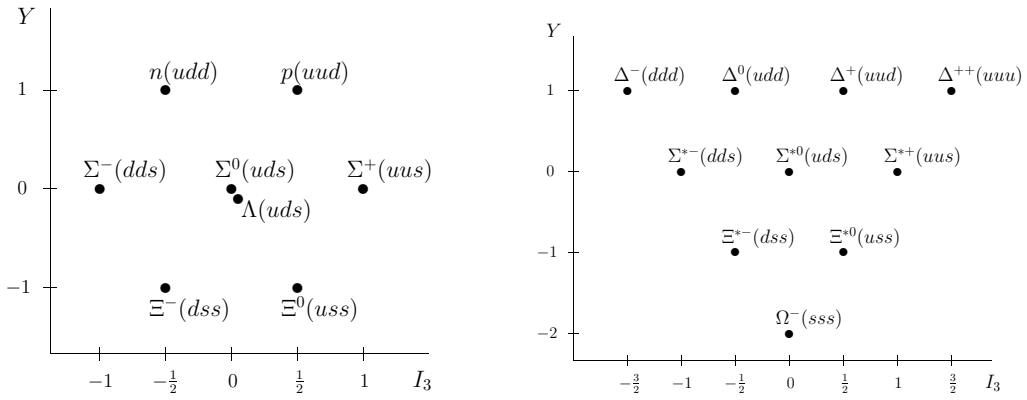

Figure 28: Quark content of octet (left) and decuplet (right) baryons.

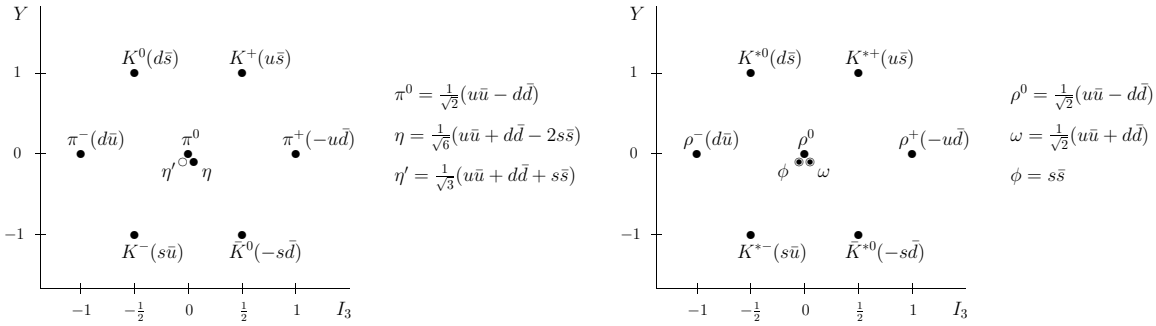

Figure 29: Quark content of octet and singlet pseudoscalar (left) and vector (right) mesons. The singlet pseudoscalar is denoted with an empty dot; the physical states coming from the mixing of the $I = 0$ octet and of the singlet states in the vector case are denoted with a half-filled dot.

and Y, together with its baryon number. Denoting with $n_{u,d,s}$ the number of times each flavour appears (minus the number of times each antiflavour appears, as usual), we have

$$I_3 = \tfrac{1}{2}(n_u - n_d)\,, \qquad Y = \tfrac{1}{3}(n_u + n_d - 2n_s)\,. \tag{5.74}$$

For one baryon we have $n_u + n_d + n_s = 3$, and imposing this constraint Eq. (5.74) can be inverted to give

$$n_u = I_3 + \tfrac{1}{2}Y + 1\,, \quad n_u = -I_3 + \tfrac{1}{2}Y + 1\,, \quad n_s = 1 - Y\,, \tag{5.75}$$

that holds for baryons. We can similarly obtain the quark/antiquark composition of each meson, solving Eq. (5.74) with the constraint $n_u + n_d + n_s = 0$. This yields

$$n_u = I_3 + \tfrac{1}{2}Y\,, \qquad n_u = -I_3 + \tfrac{1}{2}Y\,, \qquad n_s = -Y\,, \tag{5.76}$$

that holds for mesons.

Baryon number and electric charge should be assigned to quarks so that they match the experimentally observed values for the hadrons. We have already seen in Section 5.1 that this is possible; let us repeat the argument here to keep the discussion self-contained. First of all, since we know that baryon number must commute with all the SU(3) generators, we know that we can assign a baryon number to quarks. For a baryon one must have $B = B_u n_u + B_d n_d + B_s n_s = 1$

	I	I_3	Y	Q	S	B
u	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$	0	$\frac{1}{3}$
d	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{3}$	0	$\frac{1}{3}$
s	0	0	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$\frac{1}{3}$

Table 11: Quantum number assignment of quarks.

where $B_{u,d,s}$ are the baryon numbers of the various flavours, on top of $n_u + n_d + n_s = 3$, and so $0 = 3B - 3 = (3B_u - 1)n_u + (3B_d - 1)n_d + (3B_s - 1)n_s$. This must hold for all baryons, including the decuplet states uuu, ddd, and sss, where only a single flavour appears. It follows that $B_u = B_d = B_s = \frac{1}{3}$. More directly, since baryon number commutes with the SU(3) generators, it also commutes with the ladder operators that change the quark flavour, and so $B_u = B_d = B_s$; since three quarks appear in a baryon, the result above follows.

Next come electric charge and strangeness. A look at Fig. 28 shows that the proton is made of two u and one d quarks, the neutron is made of two d and one u quarks, and the Λ is made of one u, one d and one s quark. This, together with the experimental results for their charge and strangeness, and additivity of these quantum numbers, allows us to determine the value of electric charge and strangeness of each quark:

$$\begin{array}{lrcl@{\qquad}rcl} p\text{:} & 2Q_u + Q_d & = & 1\,, & 2S_u + S_d & = & 0\,, \\ n\text{:} & Q_u + 2Q_d & = & 0\,, & S_u + 2S_d & = & 0\,, \\ \Lambda\text{:} & Q_u + Q_d + Q_s & = & 0\,, & S_u + S_d + S_s & = & -1\,, \end{array} \tag{5.77}$$

from which follows $Q_u = \frac{2}{3}$, $Q_d = Q_s = -\frac{1}{3}$, and $S_u = S_d = 0$ and $S_s = -1$. Since for each flavour $f = u, d, s$ one has $Q_f = I_{3f} + \frac{1}{2}Y_f = I_{3f} + \frac{1}{2}(B_f + S_f)$, and since these are all additive quantities, the Gell-Mann–Nishijima relation and the relation between hypercharge, baryon number, and strangeness will be automatically satisfied by each baryon. As already pointed out above, antiquarks have minus the charges of quarks, so in particular baryon number, electric charge, and strangeness. For a meson made of a quark and an antiquark one will automatically find vanishing baryon number, and the Gell-Mann–Nishijima and hypercharge/strangeness relations will be satisfied again.

Finally, since the light baryons and resonances are respectively $s = \frac{1}{2}$ and $s = \frac{3}{2}$ fermions, quarks should be assigned half-integer spin $s = \frac{1}{2}$, so that only $\frac{1}{2}$ and $\frac{3}{2}$ are obtained out of the composition $\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \frac{1}{2} \oplus \frac{1}{2} \oplus \frac{3}{2}$. In general, any baryon or antibaryon will be a fermion, since its spin is determined by the composition of three half-integer quark or antiquark spins, and the two integer relative orbital momenta, and any meson will be a boson, since its spin is determined by the composition of two half-integer quark/antiquark spins, and the integer relative orbital momentum.

Explicit breaking of SU(3) So far we have been essentially working under the assumption of exact SU(3) symmetry, but we knew from the beginning that this symmetry had to be broken. Postponing a more careful treatment, here we try to assign masses to the quarks assuming that the binding energy is negligible compared to the quark masses, and based on the observation that hadronic masses grow approximately linearly with the absolute value of strangeness. We

then set for a baryon

$$\begin{aligned} m_B &= n_u m_u + n_d m_d + n_s m_s = \frac{m_u + m_d}{2}(n_u + n_d) + (m_u - m_d)\frac{n_u - n_d}{2} + n_s m_s \\ &= \frac{m_u + m_d}{2}(3 - n_s) + (m_u - m_d) I_3 + n_s m_s \\ &= \frac{3}{2}(m_u + m_d) + (-S)\left(m_s - \frac{m_u + m_d}{2}\right) + (m_u - m_d) I_3 \,. \end{aligned} \tag{5.78}$$

Within an isospin multiplet the splittings are equal to $\Delta m_B = (m_u - m_d)\Delta I_3$. Since these are known to be very small, we set in a first approximation $m_u = m_d \equiv m_{ud}$, obtaining $m_B = 3m_{ud} + (m_s - m_{ud})|S|$. From the mass splittings of particles with different strangeness we then find $m_s - m_{ud} = 150$ MeV, while from the mass of the nucleon we get $3m_{ud} \approx 940$ MeV. We then find $m_u \simeq m_d \approx 300$ MeV and $m_s \approx 450$ MeV. These masses are very different from the one discussed in the Introduction, and are referred to as *constituent masses*: the difference with the *current masses* discussed in the Introduction (see Tab. 1) comes from the fact that actually, contrary to our assumption, most of the mass of a hadron does not come from the quark masses, but rather from the interaction energy between quarks, as mediated by gluons. Furthermore, the same estimate would not work with the light pseudoscalar mesons, since linearity of masses with strangeness is not true in that case. We will return briefly on this point later on.

Quark model What we have discussed above are the basics of the *quark model* (Gell-Mann, 1964; Zweig, 1964). A few comments are in order. The SU(3) symmetry rotates quark flavours one into another, and is therefore also called flavour symmetry. It is quite far from being exact, but still quite close to it to have useful consequences. To explain the differences in baryon masses it is required to introduce an explicit breaking of the symmetry, which can be achieved for example through a different assignment of masses to the various flavours, while keeping the way they interact via strong interactions independent of the flavour. Since we know now what the microscopic theory of strong interactions is (namely, QCD), we know that this is exactly how SU(3) symmetry is broken, and so we know what kind of symmetry breaking term will appear in the strong Hamiltonian. However, the form of this term was guessed correctly before the discovery of QCD, and led to the Gell-Mann–Okubo formula to be discussed below in Section 5.2.7. Before doing that, however, we have to fix first a serious problem of the quark model.

5.2.6 Wave functions, the problem with statistics, and colour

If the quark model is to properly describe hadrons, it should be possible to assign a wave function to each baryon consistently with the fact that quarks are fermions, i.e., consistently with Fermi-Dirac statistics. More precisely, if different quark flavours are just different states of the same spin-$\frac{1}{2}$ particle, then the baryon wave functions must be totally antisymmetric under the exchange of quarks. Setting $\psi = \psi_{\rm space}\psi_{\rm spin}\psi_{\rm flavour}$, and assuming naturally that the lowest-lying states have vanishing orbital angular momenta so that the spatial wave function will be symmetric under exchange of the quarks, we have to achieve antisymmetry from the spin and flavour parts only. This can hopefully explain why only certain representations appear in nature.

Let us begin with the decuplet. Since $s = \frac{3}{2}$, the spin wave function must be symmetric. But we have seen that the flavour content of the Δ^{++} is uuu, and since the ladder operators do not

change the symmetry properties, we conclude that the flavour wave function is also symmetric for the decuplet baryons:

$$\begin{aligned}
\Delta^{++} &= |uuu\rangle\,, \\
\Delta^{+} &\propto I_-\Delta^{++} \propto |uud\rangle + |udu\rangle + |duu\rangle\,, \\
\Sigma^{*+} &\propto V_-\Delta^{++} \propto |uus\rangle + |usu\rangle + |suu\rangle\,, \\
\Sigma^{*0} &\propto V_-^2\Delta^{++} \propto |uss\rangle + |sus\rangle + |ssu\rangle\,, \\
\Omega^{-} &\propto V_-^3\Delta^{++} \propto |sss\rangle\,,
\end{aligned} \tag{5.79}$$

and similarly for the other states obtained by applying powers of I_-. This is not acceptable for fermions: if anything, our attempt to solve the representation puzzle seems to go in the wrong direction. On the other hand, everything else seems to work just fine: how can we solve this contradiction?

Before trying to deal with this, let us note that in flavour space, from the composition $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}$, we also get, besides the decuplet, a fully antisymmetric singlet,

$$\frac{1}{6}\left(|uds\rangle + |dsu\rangle + |sud\rangle - |usd\rangle - |sdu\rangle - |dus\rangle\right)\,, \tag{5.80}$$

and other $16 = 8+8$ states with mixed symmetry. To build, e.g., nucleon states, with $s = I = \frac{1}{2}$ and zero strangeness, we can start from two quarks in a $s = I = 0$ state and combine them with the remaining quark to trivially obtain $s = I = \frac{1}{2}$. The wave function of the first two quarks is then symmetric,

$$\frac{1}{\sqrt{2}}\left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\right)\frac{1}{\sqrt{2}}\left(|ud\rangle - |du\rangle\right)\,. \tag{5.81}$$

In analogy to the decuplet states, let us build a wave function which is overall symmetric in spin and flavour, and so, at first sight, unacceptable. Picking the neutron, the remaining d quark can be chosen as the first, the second or the last of the three: denoting with ψ_{ij} the antisymmetric flavour wave function of Eq. (5.81) involving quarks i and j, and by ϕ_k the flavour wave function of the k-th quark, and similarly denoting with $\tilde{\psi}_{ij}$ the antisymmetric spin wave function of Eq. (5.81) involving quarks i and j, and by $\tilde{\phi}_k$ the spin wave function of the k-th quark, we can build

$$\begin{aligned}
&\psi_{12}\phi_3\tilde{\psi}_{12}\tilde{\phi}_3 + \psi_{23}\phi_1\tilde{\psi}_{23}\tilde{\phi}_1 + \psi_{13}\phi_2\tilde{\psi}_{13}\tilde{\phi}_2 \\
&= |udd;\uparrow\downarrow\uparrow\rangle + |dud;\downarrow\uparrow\uparrow\rangle + |ddu;\uparrow\downarrow\uparrow\rangle + |udd;\uparrow\uparrow\downarrow\rangle + |dud;\uparrow\uparrow\downarrow\rangle + |ddu;\downarrow\uparrow\uparrow\rangle \\
&\quad - 2|udd;\downarrow\uparrow\uparrow\rangle - 2|dud;\uparrow\downarrow\uparrow\rangle - 2|ddu;\uparrow\uparrow\downarrow\rangle\,,
\end{aligned} \tag{5.82}$$

which is totally symmetric. Note that $\psi_{12} + \psi_{23} = \psi_{13}$: the three flavour wave functions appearing in Eq. (5.82) are not independent, and they should not be since we have only two octets in $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}$. Similarly, $\tilde{\psi}_{12} + \tilde{\psi}_{23} = \tilde{\psi}_{13}$, which again should be the case since we have only two spin-$\frac{1}{2}$ representations in the decomposition of $\frac{1}{2}\otimes\frac{1}{2}\otimes\frac{1}{2}$. Contrary to the $s = \frac{3}{2}$ case, here it is possible to construct also a totally antisymmetric flavour-spin wave function. If one were to stick to the requirements of Fermi-Dirac statistics, one would find an octet of $s = \frac{1}{2}$ baryons, but a single $s = \frac{3}{2}$ baryonic resonance, contrary to experimental evidence.

Let us now briefly turn to mesons, where we have no restriction on the symmetry of the wave function since we are combining quarks and antiquarks. Meson states can have spin $\frac{1}{2}\otimes\frac{1}{2} = 0\oplus 1$,

so the ground states with $\ell = 0$ are either $J = 0$ or $J = 1$ particles. Quantum field theory tells us that quarks and antiquarks have opposite intrinsic parity, so lowest-lying mesons are either pseudoscalars or vectors. From $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$ we obtain an octet and a singlet. As long as we deal with an exact flavour symmetry, the quark content of the pseudoscalar and vector mesons is identical. However, if we look at the quark content of physical particles, we must take into account that SU(3) is broken down to $\mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$, and so the $I = 0$ states from the SU(3) octet and the SU(3) singlet can mix. For this reason, mesons are sometimes classified into nonets, although there is no nine-dimensional irreducible representation of SU(3). It turns out that while this mixing is small for the pseudoscalars, it is almost maximal for the vectors, leading to the combinations shown in Fig. 29.

We can now go back to our antisymmetrisation problem, and discuss how that can be solved. The way out was suggested by Greenberg in 1964: add a further degree of freedom, and ask for the corresponding part of the wave function to be antisymmetric under quark exchange. To this extra degree of freedom, called *colour*, is naturally associated an extra $\mathrm{SU}(N_c)$ symmetry. Since there are no further degeneracies among hadrons masses, neither exact nor approximate, then not only the colour wave function for baryons must be antisymmetric under exchange, but in general it should also be a singlet of $\mathrm{SU}(N_c)$, for all hadrons. For N_c colours, the simplest singlet wave functions are obtained using the $\mathrm{SU}(N_c)$-invariant tensors $\delta_{i_1 i_2}$ or $\epsilon_{i_1 \ldots i_{N_c}}$, where each index runs over the values $1, \ldots, N_c$.[81] The tensor $\delta_{i_1 i_2}$ allows one to build a singlet out of the wave functions of a fundamental/complex conjugate pair, and so is appropriate for a quark/antiquark pair and so for mesons:

$$\delta_{i_1 i_2} \psi^{(1)}_{i_1} \psi^{(2)\,*}_{i_2} , \tag{5.83}$$

where $1, 2$ denote collectively the flavour and spin degrees of freedom of the quark and antiquark. The Levi-Civita tensor is instead totally antisymmetric, and allows one to build a singlet out of the wave functions of N_c fundamental objects,

$$\epsilon_{i_1 \ldots i_{N_c}} \psi^{(1)}_{i_1} \ldots \psi^{(N_c)}_{i_{N_c}} , \tag{5.84}$$

where similarly $1, \ldots, N_c$ denote collectively the flavour and spin degrees of freedom of the N_c quarks. If one assumes that there are $N_c = 3$ colours one then achieves two results at once: one obtains an "explanation" of why it takes three quarks to make a baryon; and one can combine the antisymmetric colour wave function Eq. (5.84) with a symmetric flavour/spin wave function to obtain acceptable wave functions both for the octet and the decuplet baryonic states. Moreover, since one cannot make a totally antisymmetric spin wave function out of three quarks, then one cannot use the flavour singlet wave function for baryons: the representation puzzle is then fully solved.

Although the introduction of colour was meant initially to fix the problem of the quark model with statistics, it later became the basis for the fundamental dynamical theory of strong interactions, i.e., QCD (Fritzsch, Gell-Mann, and Leutwyler, 1973). Besides using theoretical considerations to fix it, the number of colours can also be determined experimentally.

- The process $e^+e^- \to$ hadrons proceeds through annihilation of the electron-positron pair into a photon, which subsequently decays into a quark-antiquark pair. The quark and antiquark must be of the same colour type since the photon is colourless, but any colour

[81] More complicated tensors can be used as well, but let us stick to the simplest ones.

type is fine, and so the total cross section is proportional to N_c. Furthermore, if a quark-antiquark pair of flavour f is produced, then a factor Q_f^2 appears in the cross section, with Q_f the electric charge of the quark. The ratio $R = \frac{\sigma(e^+e^-\to\text{hadrons})}{\sigma(e^+e^-\to\mu^+\mu^-)}$ is then equal to $R(s) = N_c \sum_f^{\bar{N}_f(s)} Q_f^2$, with the sum running over the kinematically accessible flavours $\bar{N}_f(s)$.

- Further experimental evidence is provided by the so-called Drell-Yan process $\pi N \to \mu^+\mu^- X$, where X stands for any other possible final product besides the muon pair. This process takes place through annihilation of one of the quarks from the baryon with the antiquark in the pion into a virtual photon, which subsequently decays into a muon pair. The photon is a colour singlet, and so it can only be created if the quark and the antiquark are of the same colour type, leading to zero net colour. It there are N_c colours, only N_c out of the N_c^2 possible combinations can lead to photon production, and so the total cross section is proportional to $1/N_c$. Comparison with the cross section for $e^+e^- \to \mu^+\mu^-$ shows that $N_c = 3$.

- Another experimental confirmation comes from the neutral pion decay process $\pi^0 \to \gamma\gamma$. Here the $u\bar{u}$ or the $d\bar{d}$ component of π^0 annihilates into two photons, and each colour contributes equally to the amplitude, leading to an enhancement factor proportional to N_c^2 in the decay width, compared to the decay width for the charged-pion decay process $\pi^+ \to \mu^+\nu_\mu$.

5.2.7 The Gell-Mann–Okubo formula

As the last topic in the quark model, we discuss now how one determines the mass splittings in hadronic multiplets, i.e., how one breaks the SU(3) symmetry in order to reproduce the experimental results. With our modern knowledge of QCD and of the fact that $m_s \gg m_d \simeq m_u$, we can write down the strong interaction Hamiltonian in the quark rest frame as

$$\begin{aligned}\langle q_i|H|q_j\rangle &= M_{ij} = m_i\delta_{ij} = \text{diag}(m_{ud}, m_{ud}, m_s) \\ &= \frac{2m_{ud}+m_s}{3}\mathbf{1} + \frac{m_{ud}-m_s}{3}\text{diag}(1,1,-2) = \frac{2m_{ud}+m_s}{3}\mathbf{1} + \frac{m_{ud}-m_s}{\sqrt{3}}\lambda_8\,.\end{aligned} \tag{5.85}$$

At the static level, and at the level of quarks, we then have $H = H_0 + H_8$, where H_0 is an SU(3) singlet and H_8 transforms as the eighth component in the adjoint representation.[82] However, this kind of symmetry-breaking term had been proposed before the discovery of QCD, motivated by the fact that the breaking had to preserve both isospin and strangeness (or hypercharge, equivalently), and so had to be an isosinglet with vanishing hypercharge. The smallest representation with an element having $I = Y = 0$ is precisely the adjoint (i.e., the **8**).

In order to obtain a quantitative estimate, we now assume that H_8 is a small perturbation compared to H_0, and use the machinery of first-order perturbation theory. The ground-state baryons consist of an octet of spin-$\frac{1}{2}$ states and a decuplet of spin-$\frac{3}{2}$ states, so they cannot be mixed by a rotation-invariant perturbation, and we need worry only about the matrix elements of H_8 between members of a single irreducible representation. For each of these multiplets of

[82] This actually remains true in the full dynamical case of QCD, where the symmetry-breaking term is a mass term proportional to $\int d^3x\bar{\psi}\lambda_8\psi$; notice that this is *not* the eighth generator of the SU(3) symmetry, which is instead $\int d^3x\psi^\dagger\lambda_8\psi$.

baryons, that in the SU(3) symmetric case are exactly degenerate, i.e., $H_0|B^{(0)}\rangle = m^{(0)}|B^{(0)}\rangle$ for each unperturbed state $B^{(0)}$ in the representation, what we have to do is to diagonalise the perturbation within the degenerate subspace, i.e., diagonalise $\langle B^{(0)\prime}|H_8|B^{(0)}\rangle$. As we will now show using representation theory, the perturbation is diagonal in the basis of isospin and hypercharge eigenstates, so the masses of physical baryons are given by $m_B = m^{(0)} + \Delta m_B$ with $\Delta m_B = \langle B^{(0)}|H_8|B^{(0)}\rangle$. Moreover, representation theory allows us to determine all the Δm_B in a multiplet up to two unknown coefficients, which depend on the multiplet and on the details of the interaction.

Decomposition of an operator in irreducible components In order to see this, we will need a few pieces of information about representations in general, and about SU(3) in particular. First of all consider the matrix having as entries the matrix elements $O^{(R',R)}_{B'B} = \langle B'(R')|\mathcal{O}|B(R)\rangle$ of some operator $\mathcal{O}$ between states B and B' transforming in the representations R and R', respectively. This matrix is an object that transforms according to the representation $R \otimes \bar{R}'$, since under an SU(3) transformation

$$\begin{aligned} O^{(R',R)\prime}_{B'B} &= \langle B'(R')|\hat{U}^\dagger \mathcal{O} \hat{U}|B(R)\rangle = \sum_{\tilde{B},\tilde{B}'} (U_{R'})^*_{\tilde{B}'B'} (U_R)_{\tilde{B}B} \langle \tilde{B}'(R')|\mathcal{O}|\tilde{B}(R)\rangle = \\ &= \sum_{\tilde{B},\tilde{B}'} (U_{R'})^*_{\tilde{B}'B'} (U_R)_{\tilde{B}B} O^{(R',R)}_{\tilde{B}'\tilde{B}}\,, \end{aligned} \tag{5.86}$$

where $\hat{U}$ is the unitary operator representing the flavour transformation on the Hilbert space of states. In matrix notation

$$O^{(R',R)\prime} = U^\dagger_{R'} O^{(R',R)} U_R\,. \tag{5.87}$$

As an element of the linear space of matrices, the matrix $O^{(R',R)}$ can be written as a linear combination of basis elements, for example the matrices $e^{(R',R)B'B}$ with entries $e^{(R',R)B'B}_{\tilde{B}'\tilde{B}} = \delta_{B'\tilde{B}'}\delta_{B\tilde{B}}$. One has simply

$$O^{(R',R)} = \sum_{B',B} e^{(R',R)B'B}\, O^{(R',R)}_{B'B}\,. \tag{5.88}$$

The matrices $e^{(R',R)B'B} = e^{(R')B'} \otimes e^{(R)B}$ correspond also to the tensor product of the basis vectors $e^{(R')B'}$ and $e^{(R)B}$, that correspond in turn to the states B' and B in representations R' and R of SU(3), respectively. This shows again that $O^{(R',R)}$ transforms in the $R \otimes \bar{R}'$ representation. We know that such a direct product of representations can be decomposed in a direct sum of irreducible representations, $R \otimes \bar{R}' = \bigoplus_{\tilde{R}} \tilde{R}$: in practice, this means that the basis $\{e^{(R',R)B'B}\}$ can be traded for a new basis obtained putting together the bases of the invariant subspaces, which support the various irreducible components of $R \otimes \bar{R}'$. This can be done choosing eigenvectors of I_3 and Y, as usual, so we trade

$$\{e^{(R',R)B'B}\} \to \{\mathcal{T}^{(\tilde{R},n)}_{(R',R)j}\}\,, \tag{5.89}$$

with $\tilde{R}$ running over the irreducible representations appearing in $R \otimes \bar{R}'$ and n running over their multiplicity (a given irreducible representation may appear more than once), and j running over the weights of $\tilde{R}$ or, in other words, over the eigenvalues of I_3 and Y, as well as $\vec{I}^2$. The

matrices $\mathcal{T}^{(\tilde{R},n)}_{(R',R)j}$ obey by construction

$$U^\dagger_{R'}\mathcal{T}^{(\tilde{R},n)}_{(R',R)j}U_R = \sum_{j'}(U_{\tilde{R}})_{j'j}\mathcal{T}^{(\tilde{R},n)}_{(R',R)j'}\,. \tag{5.90}$$

After the change of basis Eq. (5.89) we then write

$$O^{(R',R)} = \sum_{\tilde{R},n,j} C^{(\tilde{R},n)}_j(R,R';\mathcal{O})\mathcal{T}^{(\tilde{R},n)}_{(R',R)j}\,, \tag{5.91}$$

with suitable coefficients $C^{(\tilde{R},n)}_i(R,R';\mathcal{O})$; in components,

$$\langle B'(R')|\mathcal{O}|B(R)\rangle = O^{(R',R)}_{B'B} = \sum_{\tilde{R},j} C^{(\tilde{R},n)}_i(R,R';\mathcal{O})(\mathcal{T}^{(\tilde{R},n)}_{(R',R)j})_{B'B}\,. \tag{5.92}$$

We can now state precisely what we mean for an operator to carry quantum numbers $I = Y = 0$: for such an operator, the only basis elements $\mathcal{T}^{(\tilde{R},n)}_{(R',R)j}$ appearing in the decomposition Eq. (5.92) are those corresponding to an isosinglet with $Y = 0$. As such, they are invariant under $\mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$ transformations, and so are suitable for breaking the SU(3) symmetry down to the symmetry actually observed in nature. Infinitely many representations contain such an element, the smallest one being the adjoint ($\mathbf{8}$) representation. Stating that an operator transforms like the eight component of an octet means that in Eq. (5.92) only terms $\mathcal{T}^{(\mathbf{8},n)}_{(R',R)I=0,Y=0}$ appear. One could similarly consider operators transforming like any element of any representation.

The discussion above generalises without much effort to any symmetry group. To see how it works in practice, consider the composition of an $s = \frac{\bar{1}}{2}$ and a $s = \frac{1}{2}$ representation of SU(2). The most general element of the tensor product of the two representation spaces is a linear superposition of $|\frac{\bar{1}}{2},\bar{s}_3\rangle \otimes |\frac{1}{2},s_3\rangle$ states, which can be conveniently written as a matrix,

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}}_{|\frac{\bar{1}}{2},-\frac{1}{2}\rangle\otimes|\frac{1}{2},\frac{1}{2}\rangle} + b \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{-|\frac{\bar{1}}{2},\frac{1}{2}\rangle\otimes|\frac{1}{2},\frac{1}{2}\rangle} + c \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{|\frac{\bar{1}}{2},-\frac{1}{2}\rangle\otimes|\frac{1}{2},-\frac{1}{2}\rangle} + d \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}_{-|\frac{\bar{1}}{2},\frac{1}{2}\rangle\otimes|\frac{1}{2},-\frac{1}{2}\rangle}\,. \tag{5.93}$$

Under an SU(2) transformation, $M \to U^\dagger M U$ with $U \in \mathrm{SU}(2)$. This matrix can also be decomposed as follows,[83]

$$\begin{aligned} M &= \frac{a+d}{2}\mathbf{1} + b\frac{\sigma_1 + i\sigma_2}{2} + c\frac{\sigma_1 - i\sigma_2}{2} + \frac{a-d}{2}\sigma_3 \\ &= \frac{a+d}{2}\mathcal{T}^{(0)}_0 + b\mathcal{T}^{(1)}_{+1} + \frac{a-d}{2}\mathcal{T}^{(1)}_0 + c\mathcal{T}^{(1)}_{-1}\,. \end{aligned} \tag{5.94}$$

The first term is clearly invariant under an SU(2) transformation, and corresponds to an $S = 0$ state. The combination of the other three terms is a traceless matrix; since the trace is invariant under SU(2) transformation, the transformed of this quantity will still be a traceless matrix, which therefore belong to an invariant subspace of dimension 3 (without further invariant

[83] For notational simplicity we drop the subscript $(\frac{\bar{1}}{2},\frac{1}{2})$ in $\mathcal{T}^{(S)}_{(\frac{\bar{1}}{2},\frac{1}{2})S_3}$.

subspaces). The matrices $\mathcal{T}^{(1)}_{1,0,-1}$ form a basis of this space, and are easily seen to be invariant up to a factor under the particular transformations $U = e^{i\alpha\frac{\sigma_3}{2}}$, i.e.,

$$U^\dagger \mathcal{T}^{(1)}_{S_3} U = S_3 \mathcal{T}^{(1)}_{S_3} \,. \tag{5.95}$$

This means that they form a triplet of $S = 1$ states with definite $S_3 = \bar{s}_3 + s_3$.

SU(3) breaking We can now focus on the case of interest. We want to break the SU(3) symmetry of the unperturbed Hamiltonian H_0 with a term that transforms like the eighth element of an octet. We write for the full strong Hamiltonian $H_{\rm strong} = H_0 + H_8$. Working in first order perturbation theory, as explained above, we need to diagonalise the matrix $\langle B'(R)|H_8|B(R)\rangle$ within the octet and the decuplet, i.e., for $R = \mathbf{8}, \mathbf{10}$. The transformation properties of H_8 means that only $\tilde{R} = \mathbf{8}$ and j corresponding to $I = Y = 0$ appear in the decomposition Eq. (5.92) for $\mathcal{O} = H_8$. It is a result in representation theory that for SU(3) and for $R' = R$, the $\mathbf{8}$ representation appears at most twice in the decomposition in irreducible representations of $\bar{\boldsymbol{R}} \otimes \boldsymbol{R}$. We can therefore write

$$\langle B'(R)|H_8|B(R)\rangle = \delta m_1(R)(\mathcal{T}^{(\mathbf{8},1)}_{(R)8})_{B'B} + \delta m_2(R)(\mathcal{T}^{(\mathbf{8},2)}_{(R)8})_{B'B} \,, \tag{5.96}$$

where $\mathcal{T}^{(\mathbf{8},i)}_{(R)8}$ are suitable matrices,[84] and $\delta m_i(R) = C^{(\mathbf{8},i)}_{I=Y=0}(R, R; H_8)$ are the corresponding coefficients in the decomposition Eq. (5.92). These multiplet-dependent coefficients cannot be determined by symmetry requirements, but they paramterise the symmetry breaking patterns within each irreducible multiplet once that we have found two suitable structures $\mathcal{T}^{(\mathbf{8},i)}_{(R)8}$ with the desired transformation properties.

The first structure is easy to find. By construction, the representatives of the generators T^a_R in representation R obey

$$[T^a_R, T^b_R] = i f_{abc} T^c_R \,, \tag{5.97}$$

from which follows

$$U^\dagger_R T^a_R U_R = (U_{\mathbf{8}})_{ab} T^b_R \,. \tag{5.98}$$

In fact, Eq. (5.97) is essentially the infinitesimal for of Eq. (5.98); the latter then follows from the former by exponentiation.[85] This means that the T^a_R transform in a $\mathbf{8}$ component within $R \otimes \bar{R}$, and so $\mathcal{T}^{(\mathbf{8},1)}_{(R)8} = T^8_R$ is the first matrix we are after.

To find the second structure we go back to the Gell-Mann matrices, and notice that the product of any two of them is necessarily of the form

$$\lambda^a \lambda^b = \frac{1}{2}\{\lambda^a, \lambda^b\} + \frac{1}{2}[\lambda^a, \lambda^b] = \frac{2}{3}\delta_{ab} + (d_{abc} + i f_{abc})\lambda^c \,. \tag{5.99}$$

[84] We dropped one of the R in the subscript for notational simplicity.

[85] Setting $T^b_R(x) \equiv e^{-ix\alpha_a T^a_R} T^b_R e^{ix\alpha_{a'} T^{a'}_R}$, one finds

$$\frac{d}{dx} T^b_R(x) = -ie^{-ix\alpha_a T^a_R}[\alpha_a T^a_R, T^b_R] e^{ix\alpha_{a'} T^{a'}_R} = i\alpha_a(-if_{abc}) e^{-ix\alpha_a T^a_R} T^c_R e^{ix\alpha_{a'} T^{a'}_R} = i(\alpha_a T^a_{\mathbf{8}})_{bc} T^c_R(x)$$

which for boundary condition $T^b_R(0) = T^b_R$ has the solution

$$T^b_R(x) = (e^{ix\alpha_a T^a_{\mathbf{8}}})_{bc} T^c_R \,.$$

It follows that $T^b_R(1) = U^\dagger_R T^b_R U_R = (e^{i\alpha_a T^a_{\mathbf{8}}})_{bc} T^c_R = (U_{\mathbf{8}})_{bc} T^c_R$.

This follows from the fact that $\lambda^a\lambda^b$ is still a 3×3 complex matrix, which can be written as a linear combination with complex coefficients of the identity matrix and of Hermitian traceless matrices, which in turn decomposes into a Hermitian part, symmetric under exchange of a and b, and an anti-Hermitian part, antisymmetric under exchange of a and b. The coefficient of the identity matrix is fixed by the normalisation of the Gell-Mann matrices, the f_{abc} are just the structure constants of the group, and the symbols

$$d_{abc} = \tfrac{1}{4}\mathrm{tr}\,\{\lambda^a,\lambda^b\}\lambda^c = 2\mathrm{tr}\,\{t^a,t^b\}t^c \tag{5.100}$$

are totally symmetric since they are invariant under cyclic permutations of the indices. An important property that follows directly form Eq. (5.100) is that d_{abc} is invariant under the transformation of all the indices via the adjoint representation,[86]

$$(U_{\mathbf{8}})_{a'a}(U_{\mathbf{8}})_{b'b}(U_{\mathbf{8}})_{c'c}d_{a'b'c'} = d_{abc}\,. \tag{5.101}$$

In turn this entails that $D^a_R \equiv d_{abc}T^b_R T^c_R$ transforms in the adjoint representation,

$$\begin{aligned} U^\dagger_R D^a_R U_R &= U^\dagger_R d_{abc}T^b_R T^c_R U_R = d_{abc}(U^\dagger_R T^b_R U_R)(U^\dagger_R T^c_R U_R) \\ &= d_{abc}(U_{\mathbf{8}})_{bb'}(U_{\mathbf{8}})_{cc'}T^{b'}_R T^{c'}_R = d_{a''bc}(U_{\mathbf{8}})_{a''a'}(U_{\mathbf{8}})_{aa'}(U_{\mathbf{8}})_{bb'}(U_{\mathbf{8}})_{cc'}T^{b'}_R T^{c'}_R \\ &= (U_{\mathbf{8}})_{aa'}d_{a'bc}T^b_R T^c_R = (U_{\mathbf{8}})_{aa'}D^{a'}_R\,, \end{aligned} \tag{5.102}$$

having used $U^{-1}_{\mathbf{8}} = U^T_{\mathbf{8}}$. In conclusion, we have for the most general symmetry breaking term with the desired transformation properties

$$\langle B'(R)|H_8|B(R)\rangle = \delta m_1(R)(T^8_R)_{B'B} + \delta m_2(R)(D^8_R)_{B'B}\,. \tag{5.103}$$

The values of d_{8bc} are explicitly known, and using them one finds

$$\begin{aligned} D^8_R = d_{8bc}T^b_R T^c_R &= -\frac{1}{2\sqrt{3}}\sum_a (T^a_R)^2 + \frac{\sqrt{3}}{2}[(T^1_R)^2 + (T^2_R)^2 + (T^3_R)^2] - \frac{1}{2\sqrt{3}}(T^8_R)^2 \\ &= -\frac{1}{2\sqrt{3}}C_R + \frac{\sqrt{3}}{2}\left(\vec{I}^{\,2} - \frac{1}{4}Y^2\right)\,. \end{aligned} \tag{5.104}$$

Here $C_R = \sum_a (T^a_R)^2$ is the quadratic *Casimir operator*, which commutes with all the generators of the group and therefore with all the elements of the irreducible representation of the group. By Schur's lemma, it must be proportional to the identity within the multiplet. More directly, since it commutes with the ladder operators and the irreducible representations are built through them starting from the highest-weight state $|\psi$, $I_+|\psi\rangle = V_+|\psi\rangle = W_+|\psi\rangle = 0$, the value of C_R is the same for all the states in the representation.[87] The remaining terms depend only on $\vec{I}^{\,2}$

[86] From the invariance of the trace and the transformation properties of t^a

$$d_{abc} = 2\mathrm{tr}\,\{U_{\mathbf{3}}t^aU^\dagger_{\mathbf{3}}, U_{\mathbf{3}}t^bU_{\mathbf{3}}\}^\dagger U_{\mathbf{3}}t^cU^\dagger_{\mathbf{3}} = (U_{\mathbf{8}})_{a'a}(U_{\mathbf{8}})_{b'b}(U_{\mathbf{8}})_{c'c}2\mathrm{tr}\,\{t^{a'},t^{b'}\}t^{c'} = (U_{\mathbf{8}})_{a'a}(U_{\mathbf{8}})_{b'b}(U_{\mathbf{8}})_{c'c}d_{a'b'c'}\,.$$

[87] One can then evaluate C_R on $|\psi\rangle$, using

$$\begin{aligned} C_R &= \vec{I}^{\,2} + \vec{V}^{\,2} - V_3^2 + \vec{W}^{\,2} - W_3^2 + \left(\tfrac{\sqrt{3}}{2}Y\right)^2 \\ &= I_3(I_3+1) + V_3 + W_3 + \left(\tfrac{\sqrt{3}}{2}Y\right)^2 + I_-I_+ + V_-V_+ + W_-W_+\,, \end{aligned} \tag{5.105}$$

and Y, and so the perturbation is diagonal in the basis of I, I_3, Y states, as it should be since we asked for $[\vec{I}, H_8] = [Y, H_8] = 0$. The diagonal terms read

$$\Delta m_B = \langle B^{(0)}|H_8|B^{(0)}\rangle = \frac{\sqrt{3}}{2}\left[\delta m_1(R) Y_B + \delta m_2(R)\left(-\frac{C_R}{3} + I_B(I_B+1) - \frac{1}{4}Y_B^2\right)\right], \quad (5.107)$$

with I_B and Y_B the total isospin and the hypercharge of baryon B. Putting everything together and redefining appropriately the unknown constants, we obtain

$$\begin{aligned} m_B &= m^{(0)}(R) + \langle B^{(0)}|H_8|B^{(0)}\rangle \\ &= \tilde{m}^{(0)}(R) + \delta\tilde{m}_1(R) Y_B + \delta\tilde{m}_2(R)\left[I_B\left(I_B+1\right) - \frac{1}{4}Y_B^2\right]. \end{aligned} \quad (5.108)$$

This is the *Gell-Mann–Okubo mass formula*. Let us check how well it works in practice.

Baryon octet For the baryon octet we have (the notation $X_{I,Y}$ is used for particle X):

$$\begin{aligned} &\Lambda_{0,0}: && m_\Lambda = \tilde{m}^{(0)} \\ &N_{\frac{1}{2},1}: && m_N = \tilde{m}^{(0)} + \delta\tilde{m}_1 + \frac{1}{2}\delta\tilde{m}_2 \\ &\Sigma_{1,0}: && m_\Sigma = \tilde{m}^{(0)} + 2\delta\tilde{m}_2 \\ &\Xi_{\frac{1}{2},-1}: && m_\Xi = \tilde{m}^{(0)} - \delta\tilde{m}_1 + \frac{1}{2}\delta\tilde{m}_2\,, \end{aligned} \quad (5.109)$$

Since there are four equations with three unknowns, one can extract one relation among masses. This can be taken to be for example

$$m_N + m_\Xi = \tfrac{3}{2}m_\Lambda + \tfrac{1}{2}m_\Sigma\,. \quad (5.110)$$

Plugging in the experimental values one finds for the LHS the value 2257 MeV, and for the RHS 2270.5 MeV, i.e., the formula is accurate to the percent level.

Baryon decuplet For the decuplet I and Y are linearly related: one immediately sees that $2I - Y$ is constant, and using, e.g., the Ω^- one finds $2I - Y = 2$. This implies

$$I(I+1) - \frac{1}{4}Y^2 = 2 + \frac{3}{2}Y\,, \quad (5.111)$$

and so the mass formula boils down to

$$m_B = \tilde{m}^{(0)} + \delta m Y\,. \quad (5.112)$$

---

and since $I_+|\psi\rangle = V_+|\psi\rangle = W_+|\psi\rangle = 0$ one finds

$$C_R = i_0(i_0+1) + \tfrac{3}{2}y_0\left(1 + \tfrac{1}{2}y_0\right) = \tfrac{4}{3}(i_0^2 + w_0^2 + i_0 w_0) + 2(i_0 + w_0)\,, \quad (5.106)$$

where $w_0 = -\frac{1}{2}i_0 + \frac{3}{4}y_0$.

This explains the very accurate linear dependence of decuplet masses on strangeness. In this case one finds

$$\begin{aligned}
\Delta_{\frac{3}{2},1}: &\qquad m_\Delta = \tilde{m}^{(0)} + \delta m\,,\\
\Sigma^*_{1,0}: &\qquad m_{\Sigma^*} = \tilde{m}^{(0)}\,,\\
\Xi^*_{\frac{1}{2},-1}: &\qquad m_{\Xi^*} = \tilde{m}^{(0)} - \delta m\,,\\
\Omega_{0,-2}: &\qquad m_\Omega = \tilde{m}^{(0)} - 2\delta m\,,
\end{aligned} \tag{5.113}$$

from which one can extract two mass relations, e.g.,

$$\begin{aligned}
m_\Delta + m_{\Xi^*} &= 2m_{\Sigma^*}\,,\\
2m_\Delta + m_\Omega &= 3m_{\Sigma^*}\,,
\end{aligned} \tag{5.114}$$

which yield LHS=2765 MeV, RHS=2768 MeV and LHS=4136 MeV, RHS=4152 MeV, respectively, i.e., an accuracy of the order of the permille. The mass of the Ω^- was initially predicted by means of this formula.

Meson octet For the pseudoscalar meson octet the Gell-Mann–Okubo formula fails disastrously. One thing we did not take into account is mixing of the flavour singlet and the flavour octet, but while this is important for the vector mesons, mixing is quite small for the pseudoscalars. However, if one uses the *square* of the masses, then the analogue of Eq. (5.110),

$$4m_K^2 = 3m_\eta^2 + m_\pi^2 \tag{5.115}$$

gives on the LHS 0.98 GeV^2 and on the RHS 0.92 GeV^2, so again a percent level of accuracy. The reason why the original formula fails can be traced back to the fact that the effect of the perturbation is not small here, but rather of the same order of the unperturbed masses: in this cases there is no reason to expect perturbation theory to work well. The reasons why the formula with the squared masses works well, instead, is hidden in the phenomenon of spontaneous breaking of chiral symmetry in QCD, a topic beyond the scope of these notes.

For completeness, we give here the derivation of the relevant coefficients d_{abc}. From Eq. (5.100) we have $d_{8bc} = \frac{1}{4}\mathrm{tr}\,\{\lambda^8, \lambda^b\}\lambda^c$. Notice that

$$\lambda^8 = \tfrac{1}{\sqrt{3}}\left(\mathbf{1}_I - 2\mathbf{0}_I\right) = \tfrac{1}{\sqrt{3}}\left(-\lambda^3 + 2\lambda^3_V\right) = \tfrac{1}{\sqrt{3}}\left(\lambda^3 + 2\lambda^3_W\right)\,, \tag{5.116}$$

where

$$\mathbf{1}_I = \begin{pmatrix} \mathbf{1}_2 & \vec{0} \\ \vec{0}^\dagger & 0 \end{pmatrix}\,, \quad \mathbf{0}_I = \begin{pmatrix} \mathbf{0}_2 & \vec{0} \\ \vec{0}^\dagger & 1 \end{pmatrix}\,, \quad \lambda^3_V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}\,, \quad \lambda^3_W = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\,. \tag{5.117}$$

The matrices $\lambda^3_{V,W}$ are the counterpart of λ^3 in the $\mathfrak{su}(2)$ subalgebras involving $\lambda^{4,5}$ and $\lambda^{6,7}$, and as such they anticommute with these matrices. One has then

$$\{\lambda^8, \lambda^{1,2,3}\} = \tfrac{2}{\sqrt{3}}\lambda^{1,2,3}\,, \quad \{\lambda^8, \lambda^{4,5}\} = -\tfrac{1}{\sqrt{3}}\{\lambda^3, \lambda^{4,5}\}\,, \quad \{\lambda^8, \lambda^{6,7}\} = \tfrac{1}{\sqrt{3}}\{\lambda^3, \lambda^{6,7}\}\,. \tag{5.118}$$

Notice furthermore that

$$\lambda^4 = \begin{pmatrix} \mathbf{0}_2 & \vec{n}_+ \\ \vec{n}_+^\dagger & 0 \end{pmatrix}\,, \quad \lambda^5 = \begin{pmatrix} \mathbf{0}_2 & -i\vec{n}_+ \\ i\vec{n}_+^\dagger & 0 \end{pmatrix}\,, \quad \lambda^6 = \begin{pmatrix} \mathbf{0}_2 & \vec{n}_- \\ \vec{n}_-^\dagger & 0 \end{pmatrix}\,, \quad \lambda^7 = \begin{pmatrix} \mathbf{0}_2 & -i\vec{n}_- \\ i\vec{n}_-^\dagger & 0 \end{pmatrix}\,, \tag{5.119}$$

where $\sigma^3\vec{n}_\pm = \pm\vec{n}_\pm$, and so

$$\left\{\lambda^3, \begin{pmatrix} \mathbf{0}_2 & \alpha\vec{n}_\pm \\ \beta\vec{n}_\pm^\dagger & 0 \end{pmatrix}\right\} = \begin{pmatrix} \mathbf{0}_2 & \alpha\sigma^3\vec{n}_\pm \\ \beta\vec{n}_\pm^\dagger\sigma^3 & 0 \end{pmatrix} = \pm \begin{pmatrix} \mathbf{0}_2 & \alpha\vec{n}_\pm \\ \beta\vec{n}_\pm^\dagger & 0 \end{pmatrix} . \tag{5.120}$$

It follows that

$$\{\lambda^8, \lambda^{4,5,6,7}\} = -\tfrac{1}{\sqrt{3}}\lambda^{4,5,6,7} . \tag{5.121}$$

Finally,

$$\{\lambda^8, \lambda^8\} = 2(\lambda^8)^2 = \tfrac{2}{3}\mathbf{1}_3 - \tfrac{1}{\sqrt{3}}\lambda^8 . \tag{5.122}$$

Since $\frac{1}{4}\mathrm{tr}\,\lambda^a\lambda^b = \frac{1}{2}\delta^{ab}$, we have in summary,

$$d_{8bc} = K_b\delta_{bc} , \qquad K_{1,2,3} = \tfrac{1}{\sqrt{3}} , \quad K_{4,5,6,7} = -\tfrac{1}{2\sqrt{3}} , \quad K_8 = -\tfrac{1}{\sqrt{3}} . \tag{5.123}$$

Colour conservation and confinement The way out of the antisymmetrisation problem for the hadronic wave functions was the introduction of the colour quantum number for quarks, with an associated colour symmetry SU(3). Colour is then a conserved quantity; one can check that it is indeed so at every vertex in QCD. On the other hand, we also required for physical states to be singlets under colour transformations, i.e., to have net colour zero. The conservation law for colour in physical processes then boils down to zero colour in, zero colour out.

The requirement for quarks and gluons is actually stronger than just being bound in colourless states: it is required of them to be permanently bound within hadrons, without the possibility of being liberated by pulling them sufficiently far apart from each other.[88] This is what is called confinement of quarks and gluons.

Although confinement is not (yet) proved within QCD, and the detailed mechanism through which it works has not yet been unveiled, nevertheless there is an argument that helps in explaining it. The (spin-independent part of the) potential between a quark and an antiquark turns out to be of the form

$$V_s = -\frac{4}{3}\frac{\alpha_s}{r} + \sigma r , \tag{5.124}$$

in the static limit of infinite quark masses. Here $\alpha_s = \frac{g_s^2}{4\pi}$ with g_s the strong coupling constant, and σ is the *string tension*. As the quark and the antiquark are pulled apart, the energy stored in the system keeps increasing linearly with the distance, and to separate them to an infinite distance (therefore liberating them) would require an infinite amount of energy. This would be strictly true if the quarks were static, i.e., with infinite mass. Since their mass is finite, for some distance R the energy stored in the system becomes sufficient to create a $q\bar{q}$ pair out of the vacuum (for $\sigma R = 2m_q$), with the new particles binding to the old ones and so preventing quark liberation. This process is called *string breaking*, and makes the potential V_s actually flat at large distances.

A Relativistic kinematics

The appropriate setting for relativistic theories is Minkowski space, in which time and space are put together in a four-dimensional entity. A point, or *event*, in Minkowski space is identified by

[88] This can change in a hot and dense environment: for sufficiently large temeperatures and/or densities, quarks and gluons get liberated from hadrons and form the so-called *quark-gluon plasma*.

four coordinates, and is denoted as X^μ with $\mu = 0, 1, 2, 3$, with X^0 the temporal coordinate and the X^j, $j = 1, 2, 3$ the spatial coordinates. In formulas,

$$X^\mu = (ct, \vec{x}) = (t, \vec{x}) \,, \tag{A.1}$$

having set the speed of light to $c = 1$. Usually, Greek indices run over $0, \ldots, 3$ and Latin indices over $1, \ldots, 3$.

Putting time and space together does not, by itself, add much to our understanding of Nature. What does, then? Let us consider first three-dimensional Euclidean space. Points in this space are identified by three coordinates as $\vec{x}$, but what makes this space the Euclidean space and not just $\mathbb{R}^3$ is how we measure distances, i.e., the choice of a *metric*. In Euclidean space distances are defined as

$$d(\vec{x}, \vec{y}) = (\vec{x} - \vec{y})^2 = (\vec{x} - \vec{y})_i (\vec{x} - \vec{y})_j \delta_{ij} \,, \tag{A.2}$$

and are clearly left invariant by translations $\vec{x} \to \vec{x} + \vec{a}$ and rotations $\vec{x} \to R\vec{x}$. Here we adopt the convention that a sum over repeated indices is understood, unless explicitly stated otherwise.

In Minkowski space, distances are replaced with the so-called *interval*,

$$\Delta s^2 \equiv (X-Y)^2 \equiv (X-Y)^\mu (X-Y)^\nu g_{\mu\nu} \equiv (X-Y)^\mu (X-Y)_\mu = (X^0 - Y^0)^2 - (\vec{X} - \vec{Y})^2 \,, \tag{A.3}$$

where $g_{\mu\nu}$ is the Minkowski metric tensor,

$$g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1) \,. \tag{A.4}$$

In Eq. (A.3) we have defined the *covariant vector* $X_\mu = g_{\mu\nu} X^\nu$, which differs from the *contravariant vector* X^μ in the sign of the spatial components,

$$X^\mu = (X^0, \vec{X}) \,, \qquad X_\mu = (X^0, -\vec{X}) \,. \tag{A.5}$$

In general, indices are lowered by $g_{\mu\nu}$, and raised by $g^{\mu\nu}$ defined by the relation $g^{\mu\rho} g_{\rho\nu} = \delta^\mu{}_\nu$. In the case at hand, $g^{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$, so as a matrix it is identical to $g_{\mu\nu}$. For future utility, we introduce the scalar product

$$X \cdot Y \equiv X^\mu Y^\nu g_{\mu\nu} = X^\mu Y_\mu = X^0 Y^0 - \vec{X} \cdot \vec{Y} \,, \tag{A.6}$$

where $\vec{X} \cdot \vec{Y}$ denotes the usual three-dimensional Euclidean scalar product. The interval is not really a distance, because it is not a positive-definite quantity. For $\Delta s^2 > 0$, we speak of a *timelike* interval; for $\Delta s^2 < 0$ of a *spacelike* interval; for $\Delta s^2 = 0$ of a *lightlike* or *null* interval. In general, for $X^2 > 0$ we speak of a timelike vector, for $X^2 < 0$ of a spacelike vector, and for $X^2 = 0$ of a lightlike or null vector.

Points Y lightlike-separated from X and such that $Y^0 - X^0 > 0$ form the *forward* or *future lightcone* of X, while those with $Y^0 - X^0 < 0$ form its *backward (past) lightcone*. Points Y timelike-separated from X and such that $Y^0 - X^0 > 0$ are inside the forward lightcone and constitute the future of event X; similarly, points Y timelike-separated from X and such that $Y^0 - X^0 < 0$ are inside the backward lightcone and constitute the past of event X. Set for simplicity $X = 0$, and consider the future of this event. An important fact is that given X_1 and X_2 inside the forward light cone, their sum $X_1 + X_2$ will still be inside the forward lightcone.

The proof is simple: first of all, notice that since $X_{1,2}$ are inside the forward lightcone, one has $(X)^2_{1,2} > 0$ and $X^0_{1,2} > 0$, which imply $X^0_{1,2} > |\vec{X}_{1,2}|$. We then have

$$\begin{aligned}(X_1 + X_2)^2 &= (X_1)^2 + (X_2)^2 + 2X_1 \cdot X_2 = (X_1)^2 + (X_2)^2 + 2(X^0_1 X^0_2 - \vec{X}_1 \cdot \vec{X}_2) \\ &> 2(X^0_1 X^0_2 - \vec{X}_1 \cdot \vec{X}_2) \geq 2(X^0_1 X^0_2 - |\vec{X}_1||\vec{X}_2|)\,,\end{aligned} \tag{A.7}$$

where we made use of Schwartz inequality. Next, we use $X^0_{1,2} > |\vec{X}_{1,2}|$ to show that $X^0_1 X^0_2 - |\vec{X}_1||\vec{X}_2| \geq 0$, and conclude

$$(X_1 + X_2)^2 > 2(X^0_1 X^0_2 - X^0_1 X^0_2) \geq 0\,, \tag{A.8}$$

i.e., $(X_1 + X_2)^2 > 0$, and obviously $X^0_1 + X^0_2 > 0$.

A.1 Lorentz transformations

In three-dimensional Euclidean space, the distance between points is invariant under rotations. The analogue in four-dimensional Minkowski space is the invariance of the interval under Lorentz transformations. These are precisely defined as the linear transformations $X' = \Lambda X$ that leave every interval invariant. The motivation behind this definition is that we want to find which transformations connecting different reference frames are consistent with Einstein's postulate that light travels at the same speed in each frame.[89] Our request is therefore

$$(X' - Y')^2 = (X - Y)^2 \Rightarrow X'^2 + Y'^2 - 2X' \cdot Y' = X^2 + Y^2 - 2X \cdot Y \Rightarrow X' \cdot Y' = X \cdot Y\,. \tag{A.9}$$

In components, $X'^\mu = \Lambda^\mu{}_\alpha X^\alpha$,

$$g_{\alpha\beta} X^\alpha Y^\beta = g_{\mu\nu} X'^\mu Y'^\nu = g_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta X^\alpha Y^\beta\,, \tag{A.10}$$

and since this must hold for all X and Y,

$$g_{\alpha\beta} = g_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta\,. \tag{A.11}$$

Using the matrix notation $\mathbf{\Lambda}_{\mu\alpha} = \Lambda^\mu{}_\alpha$, $\mathbf{g}_{\mu\nu} = g_{\mu\nu}$, Eq. (A.11) is recast as

$$\mathbf{g} = \mathbf{\Lambda}^T \mathbf{g} \mathbf{\Lambda}\,. \tag{A.12}$$

From this it follow immediately that $(\det \mathbf{\Lambda})^2 = 1$, i.e., $\det \mathbf{\Lambda} = \pm 1$, so that $\mathbf{\Lambda}$ is invertible. Transformations with $\det \mathbf{\Lambda} = 1$ are called *proper*, and leave unchanged the orientation of space (i.e., they transform a right-handed spatial coordinate system into another right-handed spatial coordinate system); *improper* transformations instead invert the orientation of space. Since clearly $\mathbf{g}^{-1} = \mathbf{g}$ is invertible,

$$\mathbf{\Lambda}^{-1} = \mathbf{g}^{-1} \mathbf{\Lambda}^T \mathbf{g}\,. \tag{A.13}$$

It is easy to see that $\mathbf{\Lambda}^{-1}$ is still a Lorentz transformation:

$$\mathbf{g} = [\mathbf{\Lambda}\mathbf{\Lambda}^{-1}]^T \mathbf{g} [\mathbf{\Lambda}\mathbf{\Lambda}^{-1}] = \mathbf{\Lambda}^{-1\,T} [\mathbf{\Lambda}^T \mathbf{g} \mathbf{\Lambda}] \mathbf{\Lambda}^{-1} = \mathbf{\Lambda}^{-1\,T} \mathbf{g} \mathbf{\Lambda}^{-1}\,. \tag{A.14}$$

[89] This request imposes Eq. (A.9) only for light-like vectors. One has to further impose the principle of relativity, i.e., equivalence of observers, to arrive at Eq. (A.10).

Let us see how this reads in component notation. From the definition of $g^{\mu\nu}$, we see that $g^{\mu\nu} = \mathbf{g}^{-1}_{\mu\nu}$. Then

$$\mathbf{\Lambda}^{-1}_{\alpha\beta} = g^{\alpha\mu}\Lambda^{\nu}{}_{\mu}g_{\nu\beta} = \Lambda_{\beta}{}^{\alpha}\,. \tag{A.15}$$

Consider now the $\alpha = 0$, $\beta = 0$ component of Eq. (A.11). We have

$$1 = \Lambda^{0}{}_{0}\Lambda^{0}{}_{0} - \Lambda^{i}{}_{0}\Lambda^{i}{}_{0}\,, \tag{A.16}$$

where sum over $i = 1, 2, 3$ is understood. Since $\Lambda^{i}{}_{0}\Lambda^{i}{}_{0} \geq 0$, we find $(\Lambda^{0}{}_{0})^2 \geq 1$, and so either $\Lambda^{0}{}_{0} \geq 1$ or $\Lambda^{0}{}_{0} \leq -1$. Transformations with $\Lambda^{0}{}_{0} \geq 1$ are called *orthocronous*. An orthocronous transformation does not change the sign of the time coordinate. In fact, consider a vector X inside the forward lightcone, and $X' = \Lambda X$. Notice first of all that from Eq. (A.11) one finds

$$g^{\alpha\beta} = g^{\mu\nu}\Lambda_{\mu}{}^{\alpha}\Lambda_{\nu}{}^{\beta}\,. \tag{A.17}$$

Using this relation for the inverse transformation $\Lambda^{-1\alpha}{}_{\beta} = \Lambda_{\beta}{}^{\alpha}$, i.e.,

$$g^{\alpha\beta} = g^{\mu\nu}\Lambda^{\alpha}{}_{\mu}\Lambda^{\beta}{}_{\nu}\,, \tag{A.18}$$

a relation analogous to Eq. (A.16) follows,

$$1 = \Lambda^{0}{}_{0}\Lambda^{0}{}_{0} - \Lambda^{0}{}_{i}\Lambda^{0}{}_{i}\,. \tag{A.19}$$

We now have for X'^{0}

$$X'^{0} = \Lambda^{0}{}_{0}X^{0} + \Lambda^{0}{}_{i}X^{i}\,. \tag{A.20}$$

The second term is the three-dimensional scalar product $\vec{\Lambda}\cdot\vec{X} \equiv \Lambda^{0}{}_{i}X^{i}$, so bounded from below by $-|\vec{\Lambda}||\vec{X}|$. But $|\vec{X}| < X^{0}$, and from Eq. (A.19) $|\vec{\Lambda}| < \Lambda^{0}{}_{0}$, so that

$$X'^{0} > \Lambda^{0}{}_{0}X^{0} - \Lambda^{0}{}_{0}X^{0} = 0\,. \tag{A.21}$$

A *proper orthocronous* Lorentz transformation therefore does not change neither the direction of time nor the orientation of space.

The proper orthocronous Lorentz transformations consist of three-dimensional rotations (the SO(3) group) and boosts. More precisely, the most general such transformation can be written as the product of a rotation times a boost in the x direction,

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.22}$$

where $\beta = \frac{v}{c} < 1$ (we work in natural units where $c = 1$), and $\gamma = 1/\sqrt{1-\beta^2}$. A boost in a general direction can obviously be obtained by first rotating x to the desired direction. The effect of the transformation Eq. (A.22) is to transform time and space coordinates t, x, y, z to new time and space coordinates t', x', y', z' according to

$$\begin{aligned} ct' &= \gamma(ct + \beta x)\,, \qquad & x' &= \gamma(x + \beta ct)\,, \\ y' &= y\,, & z' &= z\,, \end{aligned} \tag{A.23}$$

where we have reinstated the appropriate factors of c. This is recognised as the transformation relating the reference frame R, where t, x, y, z are used, to the reference frame R' (where $(t', x', y', z'$ are used) moving with speed β in the negative x direction. In fact, in the nonrelativistic limit $\beta = v/c \ll 1$ one recovers the Galilei transformations. All the other Lorentz transformations are obtained from the proper orthocronous ones by means of a *parity* transformation P or a *time reversal* transformation T,

$$P^{\mu}{}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \qquad T^{\mu}{}_{\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{A.24}$$

or by their combined action.

A.2 Point particles

Consider a point particle travelling in space. Its trajectory is described by the four-vector

$$X^{\mu}(t) = (ct, \vec{x}(t)) = (t, \vec{x}(t)), \tag{A.25}$$

where $\vec{x}(t)$ is the point in space where the particle is at time $X^0 = t$, and we have set $c = 1$. Over an infinitesimal amount of time dt, X^{μ} changes by dX^{μ},

$$dX^{\mu}(t) = (dt, d\vec{x}(t)) = dt(1, \tfrac{d\vec{x}}{dt}(t)) = dt(1, \vec{v}(t)), \tag{A.26}$$

where $\vec{v}$ is the particle velocity. It is an empirical fact that particles in Nature have $\vec{v}^2 < 1$, unless they are massless, in which case $\vec{v}^2 = 1$. The infinitesimal interval $(dX)^2$ is

$$(dX)^2 = dX^{\mu} dX_{\mu} = dt^2(1 - \vec{v}^2) \geq 0, \tag{A.27}$$

where the inequality follows from $\vec{v}^2 \leq 1$. Dividing out by dt in Eq. (A.26) one gets

$$\frac{dX^{\mu}}{dt}(t) = (1, \vec{v}(t)). \tag{A.28}$$

Despite the appearances, this quantity does not transform like X^{μ} under Lorentz transformations, i.e. $X^{\mu} \to \Lambda^{\mu}{}_{\nu} X^{\nu}$, and so it is not a Lorentz vector: in fact, while dX^{μ} is a vector, dt is not a scalar, and so dX^{μ}/dt obeys a complicated transformation law, which makes it unsuitable for a Lorentz-covariant treatment. Let us focus on the case of massive particles, for which $\vec{v}^2 < 1$ strictly. In this case it is possible to define a true vector encoding information about the particle's velocity. In this case it is in fact possible to make a Lorentz transformation to a frame in which $\vec{v} = 0$, called the *rest frame* of the particle. In this frame the spacetime trajectory reads trivially

$$X^{\mu}_{\text{rest}} = (\tau, \vec{0}), \tag{A.29}$$

where the time τ, measured in the particle's rest frame, is called *proper time*. We are taking time to flow in the same direction in the two frames, i.e., they are connected by an orthocronous Lorentz transformation. Since the interval is a relativistic invariant, we have

$$(dX_{\text{rest}})^2 = d\tau^2 = (dX)^2 = dt^2(1 - \vec{v}^2). \tag{A.30}$$

Since $dX^\mu_{\rm rest}$ is a true vector, $d\tau$ is a true scalar, independent of the reference frame. More generally, $X^2 = \tau^2$ is a Lorentz-invariant notion of time, measuring how it flows in the rest frame of a particle. Recalling that $\gamma = \frac{1}{1-\vec{v}^2}$, we find that

$$d\tau^2 = \frac{dt^2}{\gamma^2}\,. \tag{A.31}$$

Since $\gamma \geq 1$, we get the well-known time-dilation effect, i.e., $|dt| > |d\tau|$ in a frame where the particle is moving. At each instant of time, going over to the instantaneous rest frame of the particle, we can determine the amount of proper time that has elapsed for the particle between two times t_0 and t in the lab frame, where the particle moves with time-dependent speed $\vec{v}^2$:

$$\tau = \int d\tau = \int_{t_0}^{t} dt' \, \sqrt{1-\vec{v}^2(t')}\,. \tag{A.32}$$

This is the origin of the twins' paradox.

Under orthocronous Lorentz transformations, τ is an invariant, while X^μ, and similarly its differential dX^μ, transform like four-vectors, i.e. $X^\mu \to \Lambda^\mu{}_\nu X^\nu$. It then follows that the derivatives of X^μ with respect to proper time transform again like four-vectors, i.e., in the same way as X^μ does.

Using the Lorentz-invariant proper time, we can now define the four-velocity

$$u^\mu \equiv \frac{dX^\mu}{d\tau} = (\tfrac{dt}{d\tau}, \tfrac{d\vec{x}}{d\tau}) = (\gamma, \gamma\tfrac{d\vec{x}}{dt}) = (\gamma, \gamma\vec{v}) = (\gamma, \gamma\vec{\beta})\,, \tag{A.33}$$

where $\vec{\beta} = \vec{v}/c = \vec{v}$ (in natural units), which is by construction a Lorentz vector, as it is a vector divided by a scalar. Multiplying the four-velocity by the particle mass m, which is also an invariant, one obtains the four-momentum p^μ, which is again a four-vector,

$$p^\mu \equiv m u^\mu = (\gamma m, \gamma m \vec{\beta})\,. \tag{A.34}$$

Component-wise,

$$\begin{aligned} p^0 &= m\gamma = \frac{m}{\sqrt{1-\vec{v}^2}} = E\,, \\ p^i &= m\gamma\vec{\beta}^{\,i} = \frac{m\vec{v}^{\,i}}{\sqrt{1-\vec{v}^2}} = \vec{p}^{\,i}\,, \end{aligned} \tag{A.35}$$

where we have identified p^0 with the energy E and the spatial components with those of the spatial momentum $\vec{p}$. Let us check that these identifications are correct, in the sense that they reduce to the usual quantities in the non-relativistic limit. To do this, we reinstate the explicit dependence on c and take the limit $c \to \infty$, i.e., the limit of $|\vec{v}|/c \ll 1$. We have that $u^\mu = (c\gamma, c\gamma\vec{\beta})$ (check the dimensions!), and so

$$\begin{aligned} p^0 &= mc\frac{1}{\sqrt{1-\left(\frac{\vec{v}}{c}\right)^2}} = mc\left(1 + \frac{1}{2}\left(\frac{\vec{v}}{c}\right)^2 + \mathcal{O}((v/c)^4)\right)\,, \\ \vec{p} &= mc\frac{\vec{v}}{c}\frac{1}{\sqrt{1-\left(\frac{\vec{v}}{c}\right)^2}} = m\vec{v}\left(1 + \mathcal{O}((v/c)^2)\right)\,. \end{aligned} \tag{A.36}$$

The second line is, to leading order, the familiar expression for the spatial momentum of a particle. Multiplying the first line by c, we obtain

$$p^0 c = mc^2 + \frac{1}{2} m\vec{v}^2 \,, \tag{A.37}$$

which is the familiar expression for the kinetic energy of a particle, plus the *rest energy* $E_0 = mc^2$. With all the units of c in their place, then,

$$p^\mu = m\frac{dX^\mu}{d\tau} = (\tfrac{E}{c}, \vec{p}\,) \,. \tag{A.38}$$

As we have already remarked above, p^μ is a four-vector, so $p^2 = p^\mu p_\mu$ is invariant, and it is nothing but the squared mass of the particle:

$$p^2 = m^2\gamma^2(1 - \vec{\beta}^{\,2}) = m^2 \,. \tag{A.39}$$

The four-velocity square is instead simply

$$u^2 = \gamma^2(1 - \vec{\beta}^{\,2}) = 1 \,. \tag{A.40}$$

(Perhaps a more logical approach is to first introduce the four-momentum as a constant times the four-velocity, $p^\mu = mu^\mu$, then identify this constant as the particle mass using the non-relativistic limit Eq. (A.36), and conclude that m is a relativistic invariant since $p^2 = m^2$ is.) Both are timelike vectors; since $u^0 \geq 1$, both u^μ and p^μ are inside the forward lightcone. Parameterising the trajectory in terms of proper time, this means in particular that the tangent to the trajectory at $X(\tau)$ has to be within the forwards lightcone at $X(\tau)$; integrating over τ, the trajectory at times after τ has to lie within the forward lightcone at $X(\tau)$. Finally, from $p^2 = m^2$ one obtains the *dispersion relation* between energy and spatial momentum,

$$E^2 = \vec{p}^{\,2} + m^2 \,. \tag{A.41}$$

We conclude this subsection with a brief discussion of massless particles. These are particles for which the four-momentum $p^\mu = (\omega, \vec{k})$ satisfies $p^2 = 0$, i.e.,

$$\omega^2 - \vec{k}^{\,2} = 0 \Rightarrow \omega = |\vec{k}| \geq 0 \,. \tag{A.42}$$

The trajectory of these particles lies always on the lightcone.

A.3 Two-particle scattering

Consider a scattering process with two particles, a and b, in the initial state, and two particles, c and d (possibly equal to a and b), in the final state,

$$a\,b \to c\,d \,. \tag{A.43}$$

Let us discuss the kinematics of the final state both in the lab frame, in which particle b is at rest, and in the centre of mass frame, in which the total spatial momentum vanishes.

Lab frame In the lab frame the four-momenta of a and b read

$$p_a = (E_L, \vec{p}_L)\,, \qquad p_b = (m_b, 0)\,. \tag{A.44}$$

From now on we drop the coordinate index from the four-vectors. For the particles in the final state we have in general

$$p_c = (E_c, \vec{p}_c)\,, \qquad p_d = (E_d, \vec{p}_d)\,. \tag{A.45}$$

We define the angle θ_L as the angle between the trajectory of c and that of a, from the equation

$$\cos\theta_L = \frac{\vec{p}_L \cdot \vec{p}_c}{|\vec{p}_L||\vec{p}_c|}\,. \tag{A.46}$$

CM frame The CM frame is by definition the frame in which the total spatial momentum vanishes. Therefore, the four-momenta of the various particles read

$$\begin{aligned} p_a &= (E_a^*, \vec{p}^{\,*})\,, \qquad p_b = (E_b^*, -\vec{p}^{\,*})\,, \\ p_c &= (E_c^*, \vec{p}^{\,\prime *})\,, \qquad p_d = (E_d^*, -\vec{p}^{\,\prime *})\,. \end{aligned} \tag{A.47}$$

Also in this case we define the angle θ^* as the one formed by the trajectories of a and c,

$$\cos\theta^* = \frac{\vec{p}^{\,*} \cdot \vec{p}^{\,\prime *}}{|\vec{p}^{\,*}||\vec{p}^{\,\prime *}|}\,. \tag{A.48}$$

We also denote the total centre of mass energy as

$$\sqrt{s} = E_a^* + E_b^* = E_c^* + E_d^*\,. \tag{A.49}$$

We show now, using four-momentum conservation, that the energy and the magnitude of the momenta of c and d are determined uniquely in the CM, and are independent of θ^*. The values of the energies, magnitude of the momenta, and θ_L in the lab can then be obtained by means of a Lorentz transformation, and depend on the angle θ^* in the CM.

The proof reduces entirely to finding a relation between s and the individual energies of the particles. The simplest way to achieve this is to proceed as follows:

$$\begin{aligned} p_a + p_b &= p_c + p_d \\ p_b &= p_c + p_d - p_a \\ p_b^2 &= (p_c + p_d)^2 + p_a^2 - 2p_a \cdot (p_c + p_d) \\ m_b^2 &= s + m_a^2 - 2E_a^*\sqrt{s} \\ E_a^* &= \frac{s + m_a^2 - m_b^2}{2\sqrt{s}}\,. \end{aligned} \tag{A.50}$$

Notice that since s is a relativistic invariant, through this relation we can determine E_a^* from knowledge of E_L in the lab:

$$s = (p_a + p_b)^2 = m_a^2 + m_b^2 + 2p_a \cdot p_b = m_a^2 + m_b^2 + 2E_L m_b\,. \tag{A.51}$$

On the other hand, a derivation entirely analogous to that in Eq. (A.50) with a and b exchanged allows one to derive E_b^*,

$$E_b^* = \frac{s + m_b^2 - m_a^2}{2\sqrt{s}}\,, \tag{A.52}$$

and more importantly exchanging a with c and b with d we obtain the energies of the final products,

$$E_c^* = \frac{s + m_c^2 - m_d^2}{2\sqrt{s}}\,, \qquad E_d^* = \frac{s + m_d^2 - m_c^2}{2\sqrt{s}}\,. \tag{A.53}$$

From the dispersion relation we can then derive the magnitude of the momenta:

$$\begin{aligned}
|\vec{p}^{\,*}|^2 &= E_a^{*2} - m_a^2 = \frac{(s + m_a^2 - m_b^2)^2 - 4sm_a^2}{4s} = \frac{s^2 + (m_a^2 - m_b^2)^2 - 2s(m_a^2 + m_b^2)}{4s} \\
&= \frac{(s - m_a^2 - m_b^2)^2 - 4m_a^2 m_b^2}{4s} = \frac{[s - (m_a + m_b)^2][s - (m_a - m_b)^2]}{4s} \\
&= \frac{\lambda(s, m_a, m_b)}{4s}\,, \\
|\vec{p}^{\,\prime *}|^2 &= E_c^{*2} - m_c^2 = \frac{(s + m_c^2 - m_d^2)^2 - 4sm_c^2}{4s} = \frac{s^2 + (m_c^2 - m_d^2)^2 - 2s(m_c^2 + m_d^2)}{4s} \\
&= \frac{(s - m_c^2 - m_d^2)^2 - 4m_c^2 m_d^2}{4s} = \frac{[s - (m_c + m_d)^2][s - (m_c - m_d)^2]}{4s} \\
&= \frac{\lambda(s, m_c, m_d)}{4s}\,,
\end{aligned} \tag{A.54}$$

where we introduced the notation

$$\lambda(s, m_1, m_2) \equiv (s - m_1^2 - m_2^2)^2 - 4m_1^2 m_2^2 = [s - (m_1 + m_2)^2][s - (m_1 - m_2)^2]\,. \tag{A.55}$$

The magnitudes of the initial and final momenta are symmetric functions of the masses, as they should be. For completeness we briefly discuss how to recover the kinematics in the lab from that in the CM. First of all, by definition of CM, we have that

$$0 = \gamma_{\rm CM}(|\vec{p}_{\rm lab}| - \beta_{\rm CM} E_{\rm lab})\,, \tag{A.56}$$

where $\vec{p}_{\rm lab}$ and $E_{\rm lab}$ are the total spatial momentum and total energy in the lab, and so the velocity of the CM in the lab is

$$\beta_{\rm CM} = \frac{|\vec{p}_L|}{E_L + m_b}\,. \tag{A.57}$$

Writing the inverse Lorentz transformation from the CM to the lab we then find

$$\begin{aligned}
E_{c,\rm lab} &= \gamma_{\rm CM}(E_c^* + \beta_{\rm CM}|\vec{p}^{\,\prime *}|\cos\theta^*)\,, \\
|\vec{p}_{c,\rm lab}|\cos\theta_L &= \gamma_{\rm CM}(|\vec{p}^{\,\prime *}|\cos\theta^* + \beta_{\rm CM} E_c^*)\,, \\
|\vec{p}_{c,\rm lab}|\sin\theta_L &= |\vec{p}^{\,\prime *}|\sin\theta^*\,,
\end{aligned} \tag{A.58}$$

where we have made use of the fact that the transverse directions are left unaffected by a Lorentz transformation. From this and the previous relations we can obtain the kinematics of particles in the lab (the azimuthal angle transforms trivially, as it involves only transverse directions).

Example Consider a proton-antiproton ($p\bar{p}$) collision in a collider, with $E_p = E_{\bar{p}} = 270$ GeV. Clearly, $\sqrt{s} = 540$ GeV. Suppose now to perform an experiment with p at rest in the lab. What energy should the $\bar{p}$ have in the lab in order to obtain the same s? Since s is a relativistic invariant, we can evaluate it in the reference frame we prefer. In the lab

$$s = (p_p + p_{\bar{p}})^2 = 2(m_p^2 + E_L m_p) = 2m_p(m_p + E_L)\,, \tag{A.59}$$

where E_L is the energy of the $\bar{p}$ in the lab. Solving for E_L and imposing $\sqrt{s} = 540$ GeV,

$$E_L = \frac{s - 2m_p^2}{2m_p} \simeq \frac{s}{2m_p} \simeq \frac{(540)^2}{2} \text{ GeV} \simeq \frac{30}{2} \cdot 10^4 \text{ GeV} = 150 \text{ TeV}\,, \tag{A.60}$$

which is a huge energy. In general, the total CM energy scales like $E_{\rm CM} \simeq \sqrt{2m_p E_L}$.

A.4 Mandelstam variables

A convenient set of variables to describe the kinematics of $2 \to 2$ scattering processes are the so-called *Mandelstam variables*,

$$\begin{aligned} s &\equiv (p_a + p_b)^2 = (p_c + p_d)^2\,, \\ t &\equiv (p_a - p_c)^2 = (p_b - p_d)^2\,, \\ u &\equiv (p_a - p_d)^2 = (p_b - p_c)^2\,. \end{aligned} \tag{A.61}$$

The main advantage of these variables is that they are Lorentz-invariant by construction. We have already seen that s is the total centre of mass energy squared. The variable t is instead the square of the four-momentum transfer from a to c, and reads explicitly

$$t = p_a^2 + p_c^2 - 2p_a \cdot p_c = m_a^2 + m_c^2 - 2(E_a^* E_c^* - |\vec{p}^{\,*}||\vec{p}^{\,\prime *}| \cos\theta^*)\,. \tag{A.62}$$

Since energies and magnitudes of momenta are entirely determined by s and by the particle masses, we can write $t = t(s, \theta^*)$, or we can trade the variables s and θ^* used in the previous subsection with s and t, reading off $\cos\theta^*$ from Eq. (A.62). The expression for u is obtained replacing p_c with p_d, which amounts to $m_c \to m_d$ and $\cos\theta^* \to -\cos\theta^*$. It is then clear that only two of the three Mandelstam variables can be independent. In fact, one can show that

$$\begin{aligned} s + t + u &= (p_a + p_b)^2 + (p_a - p_c)^2 + (p_a - p_d)^2 \\ &= m_a^2 + m_b^2 + m_c^2 + m_d^2 + 2p_a \cdot (p_a + p_b - p_c - p_d) = m_a^2 + m_b^2 + m_c^2 + m_d^2\,. \end{aligned} \tag{A.63}$$

The Mandelstam variables satisfy various bounds that determine the physical region in which s, t, u can take values for a physical process. It is straightforward to see that

$$s \geq \max((m_a + m_b)^2, (m_c + m_d)^2)\,. \tag{A.64}$$

To get bounds on t and u, notice that

$$\begin{aligned} t = (p_a - p_c)^2 &= m_a^2 + m_c^2 - 2p_a \cdot p_c = 2(m_a^2 + m_c^2) - (m_a^2 + m_c^2 + 2p_a \cdot p_c) \\ &= 2(m_a^2 + m_c^2) - (p_a + p_c)^2 \leq 2(m_a^2 + m_c^2) - (m_a + m_c)^2 = (m_a - m_c)^2\,, \end{aligned} \tag{A.65}$$

and similarly using p_b and p_d. The same approach can be used to put an upper bound on u. We then get

$$t \leq \min((m_a - m_c)^2, (m_b - m_d)^2)\,, \qquad u \leq \min((m_a - m_d)^2, (m_b - m_c)^2)\,. \tag{A.66}$$

A lower bound is obtained combining this result and Eq. (A.63),

$$\begin{aligned} t &\geq \max(m_b^2 + m_c^2 + 2m_a m_d, m_a^2 + m_d^2 + 2m_b m_c) - s\,, \\ u &\geq \max(m_a^2 + m_c^2 + 2m_b m_d, m_b^2 + m_d^2 + 2m_a m_c) - s\,. \end{aligned} \tag{A.67}$$

Things simplify if $m_a = m_b$ and $m_c = m_d$, in which case $E_a^* = E_b^* = E_c^* = E_d^* = \frac{\sqrt{s}}{2}$, and so

$$t = m_a^2 + m_c^2 - \frac{s}{2}\left(1 - \cos\theta^* \sqrt{1 - \frac{4m_a^2}{s}}\sqrt{1 - \frac{4m_c^2}{s}}\right) . \tag{A.68}$$

Moreover, if one also has $m_a = m_c \equiv m$, then things become quite transparent, with

$$t = 2m^2 - \frac{s}{2}\left(1 - \cos\theta^* \left(1 - \frac{4m^2}{s}\right)\right) = -\left(s - 4m^2\right)\sin^2\frac{\theta^*}{2} , \tag{A.69}$$

which, since $s \geq 4m^2$, shows that

$$-\left(s - 4m^2\right) \leq t \leq 0 . \tag{A.70}$$

The upper limit is attained at threshold $s = 4m^2$ or when $\theta^* = 0$, i.e., forward scattering, while the lower limit is attained for $\theta^* = \pi$, i.e., for backscattering. This result is useful in the case of elastic processes involving only one type of particles and/or antiparticles, but also in the limit of very high energy in which we can neglect the mass in the dispersion relation and treat all particles as massless. Since in this case one simply has that $u(s, \theta^*) = t(s, \pi - \theta^*)$, the same bound applies to u; the situation at $\theta^* = 0$ and $\theta^* = \pi$ is of course the opposite of the one found for t.

Example Consider elastic pp scattering at $\sqrt{s} = 53$ GeV. The differential cross section $\frac{d\sigma}{dt}(t)$ has a peak at $-t = t_0 = 1.34$ GeV^2. What is the corresponding scattering angle in the CM?

For the elastic scattering of identical particles, $-t = (s - 4m_p^2)\sin^2\frac{\theta^*}{2} \simeq s\sin^2\frac{\theta^*}{2}$, since $s/m_p^2 \gg 1$. We have

$$\sin^2\frac{\theta^*}{2} = -\frac{t}{s - 4m_p^2} = \frac{1.34}{53^2 - 4 \cdot 0.938^2} = \frac{1.34}{2805} = 4.78 \cdot 10^{-4} . \tag{A.71}$$

To leading order, $\sin^2\frac{\theta^*}{2} \simeq \frac{(\theta^*)^2}{4}$, and

$$\theta^* \simeq 2\sqrt{5} \cdot 10^{-2} \simeq 4 \cdot 10^{-2} . \tag{A.72}$$

A convenient graphical representation of the kinematically allowed range of values of the Mandelstam variables is provided by the Mandelstam plane. Since the sum of the distances from the sides of an equilateral triangle is constant, we can take the (prolongation of the) sides of such a triangle as the $s = 0$, $t = 0$ and $u = 0$ axes, and with the right choice of size of the sides each point in the plane will be characterised by $s + t + u = m_a^2 + m_b^2 + m_c^2 + m_d^2$. In the case of equal masses, the wedge defined by the prolongation of the u and t axes identifies the physical region for the $a + b \to c + d$ process.

An important result in quantum field theory is that the scattering amplitude for the processes $a + b \to c + d$, $a + \bar{c} \to \bar{b} + d$, and $a + \bar{d} \to c + \bar{b}$, where an overbar denotes the antiparticle, are part of a single analytic function extending beyond the physical domain of the Mandelstam variables. This property is known as crossing symmetry, and relates processes in which one of the particles in the initial state is replaced with the antiparticle corresponding to one of the

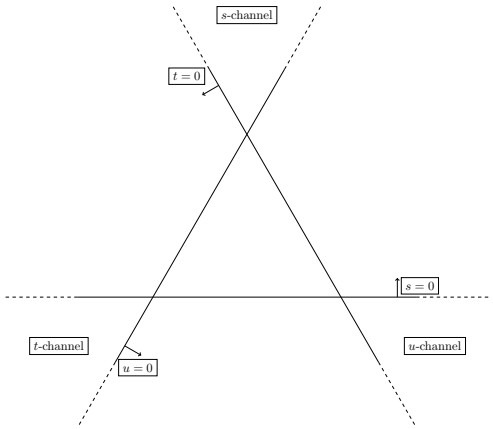

Figure 30: Mandelstam plane.

particles in the final state, and viceversa. Denoting with $A_{ab\to cd}$ the scattering amplitude for the process $a+b\to c+d$, and similarly for the other processes, one has

$$A_{ab\to cd}(p_a,p_b;p_c,p_d)=A_{a\bar{c}\to\bar{b}d}(p_a,-p_c;-p_b,p_d)=A_{a\bar{d}\to c\bar{b}}(p_a,-p_d;p_c,-p_b)\,. \tag{A.73}$$

Denoting with

$$\begin{aligned}\mathcal{A}_s(s,t,u)&=A_{ab\to cd}(p_a,p_b;p_c,p_d)\,,\\ \mathcal{A}_t(s_t,t_t,u_t)&=A_{a\bar{c}\to\bar{b}d}(p_a,p_{\bar{c}};p_{\bar{b}},p_d)\,,\\ \mathcal{A}_u(s_u,t_u,u_u)&=A_{a\bar{d}\to c\bar{b}}(p_a,p_{\bar{d}};p_c,p_{\bar{b}})\,,\end{aligned} \tag{A.74}$$

where

$$\begin{aligned}s_t&=(p_a+p_{\bar{c}})^2\,,\qquad t_t=(p_a-p_{\bar{b}})^2\,,\qquad u_t=(p_a-p_d)^2\,,\\ s_u&=(p_a+p_{\bar{d}})^2\,,\qquad t_u=(p_a-p_c)^2\,,\qquad u_u=(p_a-p_{\bar{b}})^2\,,\end{aligned} \tag{A.75}$$

the *crossing-symmetry relations* Eq. (A.73) read

$$\mathcal{A}_s(s,t,u)=\mathcal{A}_t(t,s,u)=\mathcal{A}_u(u,t,s)\,. \tag{A.76}$$

Notice that if we work with physical values of s, t and u these relations involve $\mathcal{A}_t$ and $\mathcal{A}_u$ at unphysical values of s, t, and u. In fact, working for simplicity in the case of equal masses, the physical regions of these functions are defined by $s_t\geq 4m^2, t_t\leq 0$ and $s_u\geq 4m^2, t_u\leq 0$, while in Eq. (A.75) the role of s_t and t_t is played by $t\leq 0$ and $s\geq 4m^2$, and similarly for s_u and t_u. These relations become fully meaningful if $\mathcal{A}_s$ can be analytically continued outside the physical domain, allowing to obtain, e.g., $\mathcal{A}_t$ in the corresponding physical domain from $\mathcal{A}_s$ outside the corresponding physical domain. The process $a+b\to c+d$ is called the *s-channel*, $a+\bar{c}\to\bar{b}+d$ is called the *t-channel*, and $a+\bar{d}\to c+\bar{b}$ is called the *u-channel*. The corresponding physical domains in terms of s, t and u (defined in the s-channel) are the three wedges outside the Mandelstam triangle (see Fig. 30).

A.5 Invariant phase space

The possible states of a spinless particle of mass m are characterised by the four-momenta p^μ that satisfy $p^2=m^2$ with positive energy, $p^0\geq m>0$. The corresponding domain in $\mathbb{R}^4$,

$$\Phi=\{p\in\mathbb{R}^4|p^2-m^2=0\,,\quad p^0>0\}\,, \tag{A.77}$$

is called the *phase space* of the particle. The infinitesimal element of phase space has measure

$$d\Omega_p = \frac{d^4p}{2\pi^4} 2\pi\delta(p^2 - m^2)\theta(p^0)\,. \tag{A.78}$$

This measure is manifestly invariant under orthocronous Lorentz transformations.[90] In fact, p^2 is invariant under generic Lorentz transformations, while the sign of the temporal component is invariant under the orthocronous ones.

The element of invariant one-particle phase space $d\Phi$ can be recast in a more convenient form if we make use of the general formula

$$\delta(f(x)) = \sum_{x_n, f(x_n)=0} \frac{1}{|f'(x_n)|}\delta(x - x_n)\,, \tag{A.79}$$

valid for f with simple zeros. To prove that this formula is correct, we multiply both sides by some function $h(x)$ and integrate over the real line, and show that the two sides give the same result. First, divide the real line $(-\infty, +\infty)$ into intervals I_k in which $f(x)$ is monotonic, and in each of them change variables to $y = f(x)$. Since f is monotonic in I_k it can be locally inverted, so that $x = f^{-1}(y)$ there. We get

$$\int_{-\infty}^{+\infty} dx\, \delta(f(x))h(x) = \sum_k \int_{I_k} dx\, \delta(f(x))h(x) = \sum_k \int_{f(I_k)} dy\, \frac{1}{|f'(f^{-1}(y))|}\delta(y)h(f^{-1}(y))\,. \tag{A.80}$$

Since f is monotonic in each I_k and has simple zeros, it can at most vanish once there with nonzero $|f'|$. Then only those intervals will contribute to Eq. (A.80) that contain a zero x_n, and

$$\int_{-\infty}^{+\infty} dx\, \delta(f(x))h(x) = \sum_k \int_{f(I_k)} dy\, \frac{1}{|f'(f^{-1}(0))|}\delta(y)h(f^{-1}(0)) = \sum_n \frac{1}{|f'(x_n)|}h(x_n)\,. \tag{A.81}$$

This is precisely what one obtains straightforwardly by repeating the procedure using the right-hand side of Eq. (A.79).

Using the result above, we find

$$\begin{aligned} d\Omega_p &= \frac{d^4p}{(2\pi)^3}\delta(p^2 - m^2)\theta(p^0) = \frac{d^4p}{(2\pi)^3}\delta(p^{0\,2} - \vec{p}^{\,2} - m^2) \\ &= \frac{d^4p}{(2\pi)^3}\frac{1}{2|p^0|}\left[\delta(p^0 - \varepsilon(\vec{p})) + \delta(p^0 + \varepsilon(\vec{p}))\right]\theta(p^0) = \frac{d^4p}{(2\pi)^3}\frac{1}{2\varepsilon(\vec{p})}\delta(p^0 - \varepsilon(\vec{p}))\theta(p^0) \\ &= \frac{d^3p}{(2\pi)^3 2\varepsilon(\vec{p})}\,, \end{aligned} \tag{A.82}$$

where

$$\varepsilon(\vec{p}) \equiv \sqrt{\vec{p}^{\,2} + m^2}\,. \tag{A.83}$$

[90] This remains true if we multiply $d\Omega_p$ by an arbitary numerical factor. Since the domain of integration is noncompact, the integral of $d\Phi$ is divergent, and there is no preferred choice. The one used here matches the relativistic normalisation of one-particle states, $\langle\vec{p}\,'|\vec{p}\rangle = (2\pi)^3 2p^0\delta^{(3)}(\vec{p}\,' - \vec{p})$.

The n-particle phase space is the phase space of n particles subjected to a constraint on the total four-momentum:

$$d\Phi^{(n)} = \prod_{j=1}^{n} d\Omega_{p_j} (2\pi)^4 \delta^{(4)} \left(p_{\rm tot} - \sum_{j=1}^{n} p_j \right) . \tag{A.84}$$

This is also a Lorentz invariant measure, since each $d\Phi_j$ is, and $\delta^{(4)}(\Lambda P) = |\det \Lambda|^{-1} \delta^{(4)}(P) = \delta^{(4)}(P)$.

Two-body phase space Let us work out explicitly the two-particle phase space element, which reads

$$\begin{aligned} d\Phi^{(2)} &= \frac{d^3p_1}{(2\pi)^3 2\varepsilon_1(\vec{p}_1\,)} \frac{d^3p_2}{(2\pi)^3 2\varepsilon_2(\vec{p}_2\,)} (2\pi)^4 \delta^{(4)}(p_{\rm tot} - p_1 - p_2) \\ &= \frac{1}{(2\pi)^2} \frac{d^3p_1}{2\varepsilon_1(\vec{p}_1\,)} \frac{d^3p_2}{2\varepsilon_2(\vec{p}_2\,)} \delta^{(3)}(\vec{p}_{\rm tot} - \vec{p}_1 - \vec{p}_2) \delta(E_{\rm tot} - \varepsilon_1(\vec{p}_1\,) - \varepsilon_2(\vec{p}_2\,)) \,, \end{aligned} \tag{A.85}$$

where $\varepsilon_i(\vec{p}) = \sqrt{\vec{p}^2 + m_i^2}$. We can trivially integrate over $\vec{p}_2$, setting it equal to $\vec{p}_2 = \vec{p}_{\rm tot} - \vec{p}_1$, obtaining

$$d\Phi^{(2)} = \frac{1}{(2\pi)^2} \frac{d^3p_1}{2\varepsilon_1(\vec{p}_1\,)} \frac{1}{2\varepsilon_2(\vec{p}_{\rm tot} - \vec{p}_1\,)} \delta(E_{\rm tot} - \varepsilon_1(\vec{p}_1\,) - \varepsilon_2(\vec{p}_{\rm tot} - \vec{p}_1\,)) \,. \tag{A.86}$$

We can further integrate over $|\vec{p}_1|$, so eliminating the last delta function, if we replace this by a delta function in $|\vec{p}_1|$, which requires the introduction of the appropriate Jacobian factor, as discussed above. This is most easily done using centre of mass variables, for which $\vec{p}_{\rm tot,CM} = 0$, and so $\vec{p}_{1\,\rm CM} = -\vec{p}_{2\,\rm CM}$. Let $p = |\vec{p}_{1\,\rm CM}| = |\vec{p}_{2\,\rm CM}|$. The argument of the delta function reads (with a little abuse of notation)

$$E_{\rm tot} - \varepsilon_1(p) - \varepsilon_2(p) \,, \tag{A.87}$$

and

$$\left| \frac{\partial}{\partial p} \left[E_{\rm tot} - \varepsilon_1(p) - \varepsilon_2(p) \right] \right| = \left[\frac{p}{\varepsilon_1(p)} + \frac{p}{\varepsilon_2(p)} \right] = \frac{p}{\varepsilon_1(p)\varepsilon_2(p)} \left[\varepsilon_1(p) + \varepsilon_2(p) \right] . \tag{A.88}$$

Using this in Eq. (A.86), changing variables to $d^3p_1 = dpp^2 d\Omega_{\rm CM} = dpp^2 d\cos\theta_{\rm CM} d\phi_{\rm CM}$, and integrating over p we find

$$\begin{aligned} d\Phi^{(2)} &= \frac{1}{(2\pi)^2} \frac{dpp^2 d\Omega}{2\varepsilon_1(p)} \frac{1}{2\varepsilon_2(p)} \frac{\varepsilon_1(p)\varepsilon_2(p)}{p} \left[\varepsilon_1(p) + \varepsilon_2(p) \right]^{-1} \delta(p - p_{\rm CM}) \\ &= \frac{d\Omega_{\rm CM}}{(2\pi)^2} \frac{p_{\rm CM}}{4(\varepsilon_1(p_{\rm CM}) + \varepsilon_2(p_{\rm CM}))} = \frac{d\Omega_{\rm CM}}{(2\pi)^2} \frac{p_{\rm CM}}{4E_{\rm CM,tot}} = \frac{d\Omega_{\rm CM}}{16\pi^2} \frac{p_{\rm CM}}{\sqrt{s}} \,, \end{aligned} \tag{A.89}$$

with $p_{\rm CM}$ the magnitude of the spatial momentum in the CM frame, given by Eq. (A.54). We find explicitly

$$d\Phi^{(2)} = \frac{d\Omega_{\rm CM}}{32\pi^2} \frac{\sqrt{\lambda(s, m_1, m_2)}}{s} = \frac{d\Omega_{\rm CM}}{32\pi^2} \sqrt{\left(1 - \frac{(m_1 + m_2)^2}{s} \right) \left(1 - \frac{(m_1 - m_2)^2}{s} \right)} \tag{A.90}$$

Three-body phase space Consider now the invariant phase-space element for a three-body process,

$$d\Phi^{(3)} = (2\pi)^4\delta^{(4)}(P - p_1 - p_2 - p_3)d\Omega_{p_1}d\Omega_{p_2}d\Omega_{p_3}\,. \tag{A.91}$$

Since each $d\Omega_{p_i}$ is separately invariant, we have the freedom to choose different reference frames to evaluate different parts of $d\Phi^{(3)}$, as long as the four-momenta entering the delta function are taken in the same frame. It is convenient to proceed as follows.

- We look at $d\Phi^{(3)} = d\Phi^{(3)}(P)$ as the product of the two-body phase space element $d\Phi^{(2)}(P - p_3) = (2\pi)^4\delta^{(4)}((P - p_3) - p_1 - p_2)d\Omega_{p_1}d\Omega_{p_2}$ and $d\Omega_{p_3}$.
- For particle 3 we evaluate $d\Omega_{p_3}$ in the CM of the whole system.
- For particles 1 and 2 we evaluate $d\Phi^{(2)}(P - p_3)$ in the CM of the subsystem of particles 1 and 2.

In polar coordinates we write $d^3p_3 = dp_3p_3^2d\cos\theta d\phi$. We can choose θ as the angle formed by the trajectory of particle 3 and the line of flight of the initial decaying particle or of the initial colliding particles - we will not need it for the main result. For the remaining factor we use the expression found above, Eq. (A.90), adapted to the case at hand. The total CM energy square, or invariant mass square, of the subsystem of particles 1 and 2 is $s_{12} = (p_1+p_2)^2$, and the spatial momentum in the corresponding CM frame has magnitude $p'^2_{\rm CM} = \lambda(s_{12}, m_1, m_2)$ (we will not need the detailed expression). Here and below the prime is used to distinguish quantities in this frame from the unprimed ones evaluated in the CM of the whole system. For the angular variable θ' we choose the angle between the trajectory of particles 1 and 3 in the CM frame of particles 1 and 2. Putting everything together we find

$$d\Phi^{(3)} = \frac{dp_3p_3^2}{E_3}\frac{d\cos\theta d\varphi}{2(2\pi)^3}\frac{d\cos\theta' d\varphi'}{16\pi^2}\frac{p'}{\sqrt{s_{12}}}\,, \tag{A.92}$$

with E_3 the energy of particle 3 in the CM frame, and $p' = |\vec{p}_1'| = |\vec{p}_2'|$ the magnitude of the momenta of particles 1 and 2 in the corresponding CM frame. This can be expressed in a much more useful form if we trade p_3 for s_{12}, and $\cos\theta'$ for $s_{13} = (p_1 + p_3)^2$, i.e., the total CM energy square of the subsystem of particles 1 and 3. For the first step we change variables from p_3 to E_3, which since $E_3dE_3 = p_3dp_3$ leads to

$$\frac{dp_3p_3^2}{E_3} = \frac{dE_3E_3p_3}{E_3} = dE_3p_3\,, \tag{A.93}$$

and use the relations

$$\begin{aligned} E_3 &= \frac{s + m_3^2 - s_{12}}{2\sqrt{s}}\,, \\ p_3^2 &= \frac{\lambda(s, m_3, \sqrt{s_{12}})}{4s} = \frac{(s - m_3^2 - s_{12})^2 - 4m_3^2s_{12}}{4s}\,, \end{aligned} \tag{A.94}$$

that follow from Eqs. (A.50) and (A.54) if we treat the 1-2 subsystem as a single particle. For the second step we evaluate s and s_{13} in the CM of the 1-2 subsystem,

$$\begin{aligned} s &= (p_1 + p_2)^2 + m_3^2 + 2(p_1 + p_2)\cdot p_3 = s_{12} + m_3^2 + 2\sqrt{s_{12}}E_3'\,, \\ s_{13} &= m_1^2 + m_3^2 + 2(E_1'E_3' - |\vec{p}_1'||\vec{p}_3'|\cos\theta') = m_1^2 + m_3^2 + 2(E_1'E_3' - p'p_3'\cos\theta')\,, \end{aligned} \tag{A.95}$$

From the first relation and the mass-shell condition we obtain

$$\begin{aligned} E_3' &= \frac{s - m_3^2 - s_{12}}{2\sqrt{s_{12}}}\,, \\ p_3'^2 &= \frac{(s - m_3^2 - s_{12})^2 - 4 s_{12} m_3^2}{4 s_{12}} = \frac{s}{s_{12}} p_3^2\,. \end{aligned} \tag{A.96}$$

Taking the differential of the first relation in Eq. (A.94) and of the second relation in Eq. (A.95) we find

$$\begin{aligned} dE_3 &= -\frac{ds_{12}}{2\sqrt{s}}\,, \\ ds_{13} &= -2 p' p_3' d\cos\theta'\,. \end{aligned} \tag{A.97}$$

Plugging Eqs. (A.93), (A.94), (A.96) and (A.97) into (A.92) we finally obtain

$$\begin{aligned} d\Phi^{(3)} &= \frac{ds_{12}\, p_3}{2\sqrt{s}} \frac{ds_{13}}{2 p' p_3'} \frac{p'}{\sqrt{s_{12}}} \frac{d\cos\theta d\varphi d\varphi'}{2(2\pi)^3 16\pi^2} = \frac{ds_{12}\, p_3}{\sqrt{s}} \frac{ds_{13}\,\sqrt{s_{12}}}{p_3 \sqrt{s}} \frac{1}{\sqrt{s_{12}}} \frac{d\cos\theta d\varphi d\varphi'}{8(2\pi)^3 16\pi^2} \\ &= \frac{ds_{12}\, ds_{13} d\cos\theta d\varphi d\varphi'}{(4\pi)^5 s}\,. \end{aligned} \tag{A.98}$$

The most important aspect of this formula is that the three-body phase-space element is independent of the invariant masses squared of the 1-2 and 1-3 subsystems, as well as of the remaining angular variables. This means that if we measure s_{12} and s_{13} in a process with three particles in the final state, and for each event we put a dot in the (s_{12}, s_{13}) plane, then the density of dots is not affected by any kinematical factor, and reflects dynamical features only. A plot of this type is known as *Dalitz plot*. In other words, if the probability of obtaining any pair of possible values (s_{12}, s_{13}) is flat, i.e., independent of s_{12} and s_{13}, then the Dalitz plot would show a uniform density of dots within the boundaries of the allowed region.[91] If, on the other hand, certain values are preferred, e.g., because the process takes place mostly through an intermediate two-body state with a subsequent two-body decay of an unstable particle, then there would be regions of the Dalitz plot where dots are denser, and one could identify the masses of the resonant intermediate state.

[91] These boundaries are determined by the maximal and minimal possible values of s_{13} as a function of s_{12}, obtained by setting $\cos\theta' = \mp 1$ in Eq. (A.95), and using Eq. (A.96) to express E_3' and p_3' as functions of s_{12}.

References

[1] David J. Griffiths, *Introduction to elementary particles*, 2nd edition (Wiley, New York, 2008).

[2] P. A. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020**, 083C01 (2020) and 2021 update

[3] J. R. Taylor, *Scattering theory: the quantum theory of nonrelativistic collisions*, (Wiley, New York, 1972).

[4] T. D. Lee and C. N. Yang, Phys. Rev. **104** (1956), 254-258

[5] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, Phys. Rev. **105** (1957), 1413-1414

[6] R. L. Garwin, L. M. Lederman and M. Weinrich, Phys. Rev. **105** (1957), 1415-1417

[7] Steven Weinberg, *The quantum theory of fields. Vol. 1: Foundations* (Cambridge University Press, Cambridge, 1995).

[8] Valentine Bargmann, Ann. Math. **59** (1954), 1-46.

[9] Morton Hamermesh, *Group theory and its application to physical problems*, (Addinson-Wesley, Reading, 1962).

[10] Sidney Coleman, *Aspects of Symmetry: Selected Erice Lectures*, (Cambridge University Press, Cambridge, 1985).

[11] Otto Nachtmann, *Elementary Particle Physics: Concepts and Phenomena*, (Springer, Berlin, Heidelberg, 1990).

[12] A. Pevsner *et al.*, Phys. Rev. Lett. **7** (1961), 421-423.

[13] V. E. Barnes *et al.*, Phys. Rev. Lett. **12** (1964), 204-206.

[14] M. Gell-Mann, "The Eightfold Way: A theory of strong interaction symmetry", California Institute of Technology Synchroton Laboratory Report CTSL-20 (1961) (unpublished); Y. Ne'eman, Nucl. Phys. **26** (1961), 222