Notes on elementary particle physics

Matteo Giordano

December 6, 2018

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, if you wish. 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 lead 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 "elementary" 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.

1.1 Brief history of particle physics

The birth of elementary particle physics can be considered the discovery of the *electron* (e) by Thomson in 1897. Making experiments with cathode rays he noticed that they were deflected by a magnetic field, which hinted at them actually being streams of light, negatively charged particles (from now on, the absolute value of the electron charge will define the unit of electric charge). He 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, and he estimated their energy from the heat generated by the stream when hitting the anode. 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 Rutherford (also making use of the experimental results of Geiger and 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. He later understood that each unit of positive charge in a nucleus corresponded to one hydrogen nucleus, which came to be known as *proton* (p). There was still a mismatch between the mass of the nucleus and what would have been expected from its charge: this was settled by the discovery of the *neutron* (n) by Chadwick in 1932. The neutron forms part of the nuclei, has almost the same mass as the proton, but is electrically neutral.

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 (Einstein, 1905) and the scattering of light on particles at rest (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.

It still remained to be explained how protons could stay together in nuclei despite the electric repulsion (and how neutrons could be bound there, too). This 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, Yukawa proposed in 1934 that the strong interaction was

[1] This does not apply to wearing socks with sandals: that is and always will be an abomination which deserves a special place in Hell.

[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.

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"). Toghether, baryons and mesons form the class of strongly interacting particles, the *hadrons* (from the Greek "strong"). 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 (Anderson & Neddermeyer, 1936)... except it was *not* Yukawa's meson! More detailed studies of cosmic rays in 1946-47 (Powell, Lattes, Occhialini, 1947) led to conclude that there were *two* types of particles in the cosmic rays: 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 the nuclei. The first particle is the *pion* (π), while the second one is the muon (μ), into which the pion decays. While the pion was the missing piece in the strong interaction puzzle, the muon was not expected (and 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 "light").

In the meantime, it had become clear that each particle had a "twin", called *antiparticle*, with the same mass but opposite electrical charge. In fact, after having been predicted theoretically by Dirac, in 1931 the antielectron, or *positron*, had been observed experimentally by 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[3] were also observed, by Segré and Chamberlain, and by Cork, respectively.

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 had to be neutral and (at least almost) massless. This led to the *neutrino* (ν) hypothesis proposed by Pauli in 1930. Direct observation of neutrinos proved to be extremely difficult, since these particles interact very weakly, and it had to wait until the mid-fifties (Reines & 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*, had still to be properly understood, but we will return on this later.

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: this was the kaon (K), the first of what will become known as "strange" particles. 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 Gell-Mann and (independently) Zweig proposed the *quark model* in 1964. In this model, mesons and baryons are not elementary, but instead bound states of quarks and antiquarks,

[3]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.

leptons				quarks			
particle	charge	mass	spin	particle	charge	mass(*)	spin
e^-	-1	0.511 MeV	$\frac{1}{2}$	u	$\frac{2}{3}$	2.2 MeV	$\frac{1}{2}$
μ^-	-1	105.7 MeV	$\frac{1}{2}$	d	$-\frac{1}{3}$	4.7 MeV	$\frac{1}{2}$
τ^-	-1	1.777 GeV	$\frac{1}{2}$	s	$-\frac{1}{3}$	96 MeV	$\frac{1}{2}$
ν_e	0	< 2.2 eV	$\frac{1}{2}$	c	$\frac{2}{3}$	1.28 GeV	$\frac{1}{2}$
ν_μ	0	< 0.17 MeV	$\frac{1}{2}$	b	$-\frac{1}{3}$	4.18 GeV	$\frac{1}{2}$
ν_τ	0	< 15.5 MeV	$\frac{1}{2}$	t	$\frac{2}{3}$	173.1 GeV	$\frac{1}{2}$

Table 1: Matter particles. 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.[5]

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). 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 in 1974, the *charm* (c).

From here on, a further lepton, the tau (τ) was discovered in 1975 by Perl;[4] the corresponding neutrino was immediately theorised, but observed only in 2000 by the DONUT experiment at Fermilab. A further quark was observed in 1977 by Lederman (see footnote 4), the *bottom* or *beauty* (b), which led to theorise a sixth one, the *top* or *truth* (t), finally observed in 1995 by the CDF and DØ experiments at Fermilab.

To complete the story, the analogues of the photon in the weak and strong interactions have to be identified. In fact, the success of Quantum Electrodynamics (QED), based on the interpretation of the photon as the mediator of the electromagnetic interaction, stimulated attempts at modeling the weak and strong interactions in a similar way. Moreover, the success of the quark model 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); the theory describing the interaction of quarks and gluons is known as Quantum Chromodynamics (QCD), for reasons to be explained later. As for the weak interactions, their mediators are the *intermediate vector bosons*, namely the W (charged) and Z (neutral) bosons, experimentally observed in 1983 by the UA1 and UA2 experiments at CERN. The last brick is the so-called *Higgs boson* (H), observed in 2012 by the ATLAS and CMS experiments at CERN, whose role will be discussed later. The theory describing weak interactions, and also unifying them with electromagnetism, is known as the Glashow-Salam-Weinberg model. Together with QCD, it forms the *Standard Model* of particle physics. The hypothetical quantum of gravitational interactions, the *graviton* (G) has not been experimentally observed yet.

[4] This should be understood in the sense of Perl *and collaborators*: the era of solitary discoveries in particle phyisics was long gone by then.

[5] 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.

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

Table 2: Interaction particles.

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 elementary particles span five orders of magnitude in mass, from the 0.5 MeV of the electron to the 170 GeV of the top quark.

1.2.1 Matter particles

Matter particles have a common feature: they are all spin-$\frac{1}{2}$ fermions.[6] 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 possess electrical charge, and the corresponding electronic, muonic and tauonic *neutrinos*, which are electrically neutral. To each particle correspond 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* (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, to 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:

[6] 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 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.

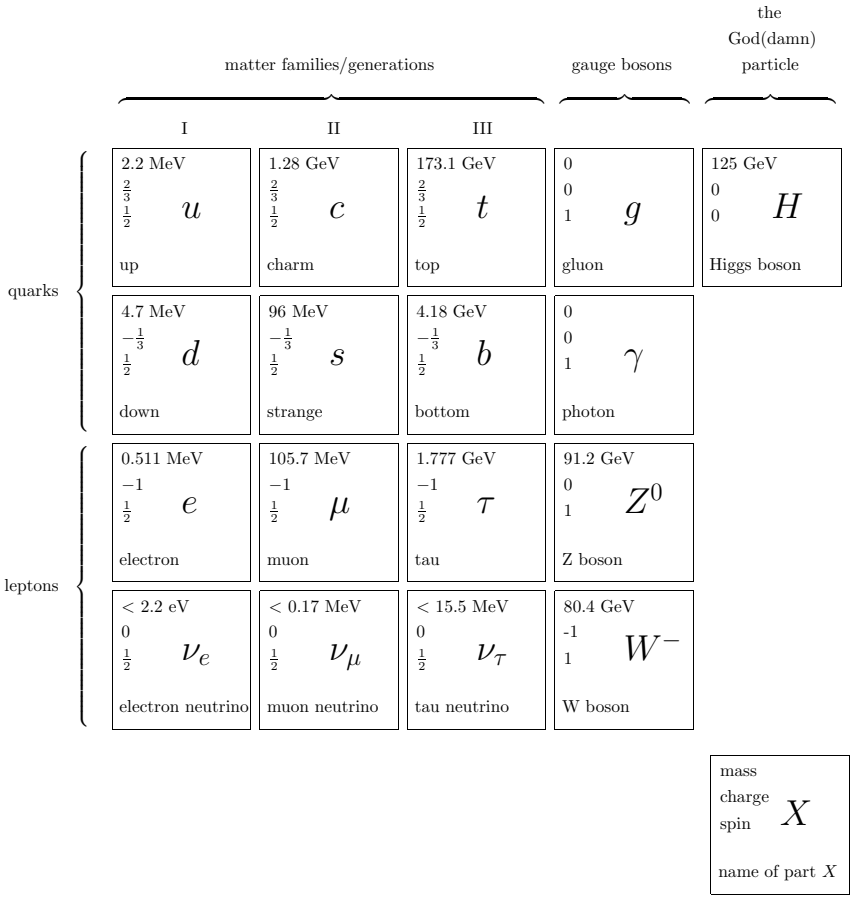

Figure 1: "Periodic table" of elementary particles (the existence of antiparticles is understood).

- 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 among 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, and which essentially provides the mass to all other elementary particles. It interacts with the intermediate vector bosons but not with the photon or with the gluons (hence these are massless).

1.3 Natural units

You may have noticed that I have expressed the masses of particles using energy units. This is possible because of the existence of a fundamental constant, the speed of light c, that allows to translate masses into energies. The correct mass unit should therefore be $[m] = \mathrm{eV}/c^2$ (recall the Einstein relation $E = mc^2$). Similarly, the existence of the Planck constant $\hbar$ allows to express times in terms of inverse energies, i.e., in units of $[t] = \hbar/\mathrm{eV}$ (recall here the relation between the energy of a photon and its frequency, $E = \hbar\nu$). Using together c and $\hbar$ we can express lengths in units of $[l] = c\hbar/\mathrm{eV}$. Combining this with the relation between energy and mass we can write a length in units of $c\hbar/[c^2(\mathrm{eV}/c^2)] = \hbar/[c(\mathrm{eV}/c^2)] = \hbar/(c[m])$. Finally, from the Coulomb potential energy (in Heaviside-Lorentz units) $U = e^2/(4\pi r)$ we find $[e]^2 = [E][l] = [\hbar c]$. Notice that the two constants originate respectively from special relativity and quantum mechanics. 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 to a particle is the Compton length $\lambda_C = \frac{\hbar}{mc}$; 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 $e^2/(4\pi)$.

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

1.4 Building up matter

From the elementary particles in Table 1, via the interactions mediated by the particles in Table 2, one can ultimately build up all the matter surrounding us, all the way from the proton to János bácsi and Viki néni. This happens in several stages, characterised by different length and energy scales.

The first stage is the construction of hadrons from quarks. It is a fact of life that free quarks are not observed in Nature, but always come along bounded into mesons and baryons. This phenomenon is known as *confinement*. Mesons are essentially bound states of a quark and an antiquark, while baryons are made up of three quarks.

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 a d quark, followed by the neutron with two d and one u quark. 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),

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

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

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	$\pi^0 \to \gamma\gamma$	$8.4 \cdot 10^{-17} s$

Table 4: Decays of unstable particles.

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 center of the hydrogen atom, or together with other protons and neutrons[7] 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. As with the mesons, heavier baryons exist with the same quark content but in different 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.

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. Different kinds of atoms have different amounts of protons in their nuclei, and can occur with a different number of

[7]Neutrons can be stable in nuclei, although not in all of them.

neutrons (isotopes). The rest of the atom consists of electrons, bound to the nucleus by the electromagnetic interactions. The electron is the lightest charged particle and is stable (again, luckily for us) thanks to conservation of electric charge. Electromagnetic interactions are also 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.

1.5 Unstable particles and decays

The proton, the electron, the neutrinos,[8] the photon and their corresponding antiparticles, are the only stable particles.[9] Other hadrons and leptons, as well as the intermediate vector bosons,[10] *decay*, i.e., they "break up" in various ways 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 the lifetimes of these, is given in Table 4.

The meaning of lifetime is related to the empirical (approximate, but rather accurate) finding that the decay rate (i.e., the probability of decaying per unit time per particle) of unstable particles is independent of time (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 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

$$dN(t) = -\Gamma N(t) dt \,, \tag{1.1}$$

and so one finds the exponential decay law

$$N(t) = N(0) e^{-\Gamma t} = N(0) e^{-\frac{t}{\tau}} \,, \tag{1.2}$$

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

$$M^2 = P^2 = P \cdot \sum_i p_i = M \sum_i E_i^{CM} \geq M \sum_i m_i \,, \tag{1.3}$$

[8] 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).

[9] 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.

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

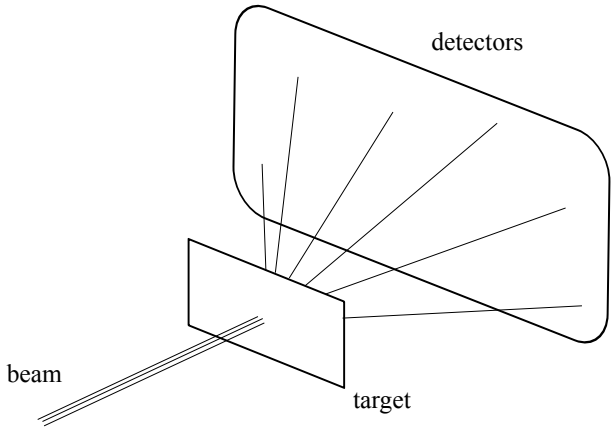

Figure 2: Setup of a fixed target experiment.

i.e., 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, which can lead to conservation laws. Examples are conservation of electric charge, baryon number, lepton and (approximately) lepton family number (see below).

1.6 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 itself) was created, and their lifetime is long enough, then part of them might have survived this long and can be still observed. This is the case, e.g., for the uranium nuclei. However, this does not really answer our question: how can have they been created in the first place?

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 phyisics. In this subsection I will just introduce the basic concepts, focussing on so-called *fixed-target* experiments.

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. 2). 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 particles that it contains, by their velocity v, and by its cross-sectional area A_b. The target is characterised

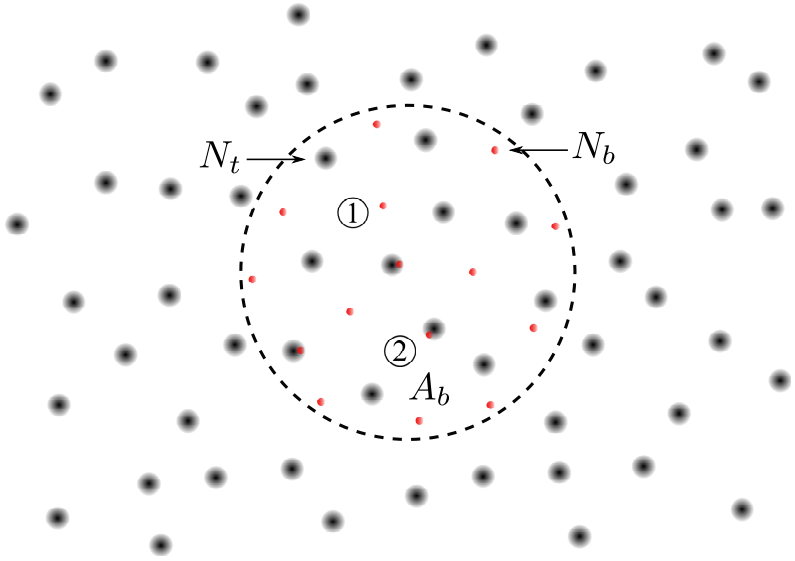

Figure 3: The target as seen from the beam. Target particles are the black discs, with area corresponding to the cross section σ of the process, beam particles are the red dots. The dashed line indicates the cross-section of the beam, of area A_b, N_b is the number of particles in the beam, and N_t is the number of particles in the target "seen" by the beam. The beam particle 1 does not undergo scattering, while particle 2 does.

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 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. From the point of view of the beam the target looks like a two-dimensional surface with "active" points corresponding to the scatterers, so practically the target can be imagined as being collapsed on a two-dimensional sheet (see Fig. 3). The area of this region 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. The proportionality factor is the ratio between the area of the "good part" of the target surface (corresponding to the "active" area around a target particle) where the beam particle has to impinge, and the cross-sectional area of the beam, i.e., the probability that one given particle in the beam (that we are assuming can be found anywhere on its cross-section with uniform probability) happens to pass nearby one given particle in the target. If the target is thin enough, a particle in the beam will not undergo multiple scattering processes.

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 the energy of the beam particles). The cross section has dimensions of an area, and in fact measures the area of the "active" region around the target discussed above. In formulas,

$$N_{\text{events}} = N_t N_b \frac{\sigma}{A_b}\,, \tag{1.4}$$

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

$$\sigma = \frac{N_{\text{events}}}{N_t \frac{N_b}{A_b}}\,. \tag{1.5}$$

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 simplicitly 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. (1.4) with N_b replaced by ΔN_b, and so in Eq. (1.5) 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

$$\sigma = \frac{\Delta N_{\text{events}}}{\Delta t N_t \Phi}\,. \tag{1.6}$$

i.e., the cross section is the number of scattering event per unit time, unit target and unit flux.

Both scattering and decay processes are governed by the fundamental interactions, which tell us which processes can take place, and with what probability. The considerations 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 first at how the various interactions work.

Exercise When scattering on a target particle, a projectile is either deflected (elastic scattering) or involved in some inelastic process: in any case, it will be removed from the beam. Consider a beam of N_b particles directed at a target of thickness d and density $N_t/V = \rho_t$, and let the total scattering cross section be σ. How many particles are expected to emerge on the other side of the target? [Answer: $N_b e^{-\frac{d}{\lambda}}$ with $\frac{1}{\lambda} = \sigma\rho_t$.]

1.7 Describing interactions: Feynman diagrams

The simplest way to describe how matter particles interact with each other is to look at interactions as the exchange of the mediators of the various forces between them. The fundamental process consists therefore in a matter particle emitting or absorbing an interaction particle. If the interacting particle is then absorbed by a second matter particle, or had been emitted by that, then it mediates the interaction between the two particles. Actually it is not really well

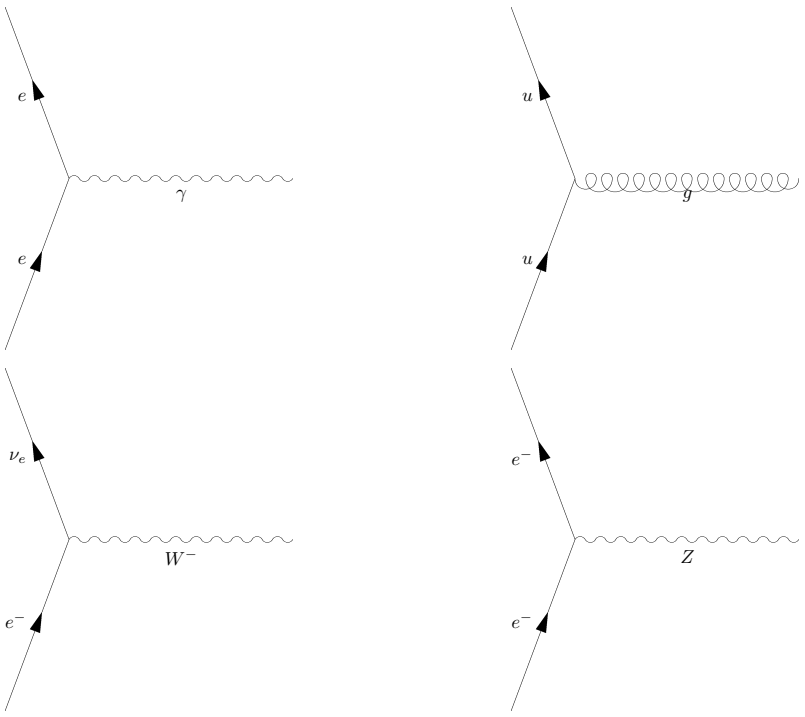

Figure 4: The fundamental vertices involving matter particles. Top left: electromagnetic vertex. Top right: strong interaction vertex. Bottom left: weak interaction vertex, charged current. Bottom right: weak interaction vertex, neutral current.

defined which particle is the emitter and which is the absorber: we are dealing with very short time scales and we cannot really observe the exchange process. Luckily enough, this does not matter: all that matters is that an interaction particle has been exchanged.

The fundamental process described above is called an interaction *vertex*. All the quantities that are conserved at a vertex will be automatically conserved by the interaction: these include for example energy and momenta, angular momentum, and the electric charge, which are conserved by all interactions. Other quantities may or may not be conserved at a certain type of vertex.

Let us begin with the electromagnetic interactions. The fundamental vertex is shown in the top left panel of Fig. 4: an electron (but it could be any other charged lepton or a quark) enters, emits or absorbs a photon, and exits. Time is assumed to flow upwards. This is the only basic process in QED. A diagram like this is known as *Feynman diagram*, and there is much more to it than simply a pictorial representation of a process - as we will learn in due time.

For strong interactions a similar process occurs, with a quark emitting or absorbing a gluon, without changing its flavour. This is shown in the top right panel of Fig. 4. The main difference is that quarks and gluons carry an extra degree of freedom, called *colour*, and the emission/absorption of a gluon can change the colour of the quark. Overall, though, colour is conserved at the vertex, as the difference (if any) is carried away by the gluon. Quarks come in three colours, and there are eight type of gluons (of the nine possible colour combinations, that corresponding to gluons that leave all the colours unchanged does not appear). Besides this process there are two other processes in QCD involving only gluons (Fig. 5).

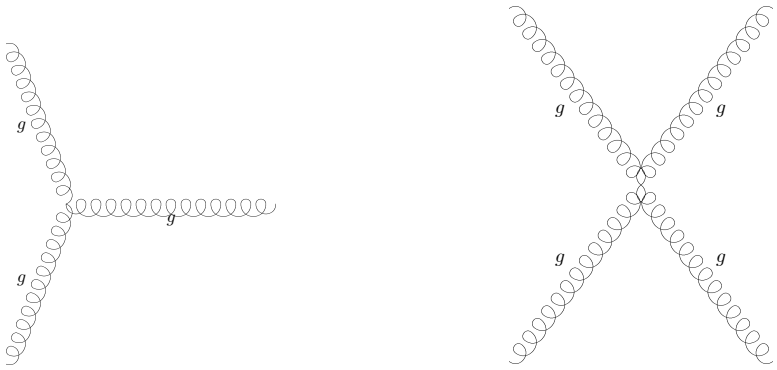

Figure 5: Interaction vertices involving only gluons.

Weak interactions between matter particles proceed through two possible processes: an electron emits/absorbs a W^- and turns into an electronic neutrino (lower left panel of Fig. 4); or it emits/absorb a Z and exits (lower right panel of Fig. 4). The first process involves a charged current, while the second one involves a neutral current. Charged currents involve leptons in the same family, so a muon will turn into a muonic neutrino and a tau will turn into a tauonic neutrino. There is clearly another set of possible processes, that in which an antilepton emits a W^+ and turns into an antineutrino of the appropriate kind, but this is just the antiparticle version of the first process (see below). Also in this case there are other vertices not involving matter particles, but only intermediate vector bosons and photons (Fig. 6).

Vertices involving antiparticles are constructed by drawing an antifermion travelling forward in time like a fermion travelling backwards in time. This corresponds to reflecting the vertices of Fig. 4 vertically. Look at Fig. 7 if this sounds confusing. The vertices of Figs. 4 and 7 can be "rotated" leading to Fig. 8: here a fermion and an antifermion enter the process.

The lepton-lepton-boson vertices of Figs. 4, and their reflected and rotated versions in Figs. 7 and 8, allow to deduce an important conservation law. In Fig. 4 a lepton enters and a lepton exits; in Fig. 7 an antilepton enters and an antilepton exits; in Fig. 8 a lepton and an antilepton enter, and no lepton/antilepton exits. Since electrons couple through the charged weak interaction only to electronic neutrinos, and similarly for the muon and the tau, the total number of lepton of a given family is conserved separately. More precisely, if we assign *family lepton numbers* $L_{e,\mu,\tau} = 1$ 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, then the total $L_{e,\mu,\tau}$ is the same before and after the process. Nowadays it is known that the total family lepton numbers are actually not exactly conserved. Indeed, this would be the case if neutrinos were exactly massless. For nonzero neutrino masses, it is possible for them to "oscillate" changing their type, e.g., from electronic to tauonic, thus violating family lepton number conservation. Such oscillations have actually been observed. On the other hand, the total *lepton number* $L = L_e + L_\mu + L_\tau$ remains conserved also in the presence of neutrino oscillations.

We have not discussed yet weak vertices involving quarks. These are almost exactly the same as those involving leptons, but there is a twist. The process involving the weak neutral current works exactly in the same way. 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

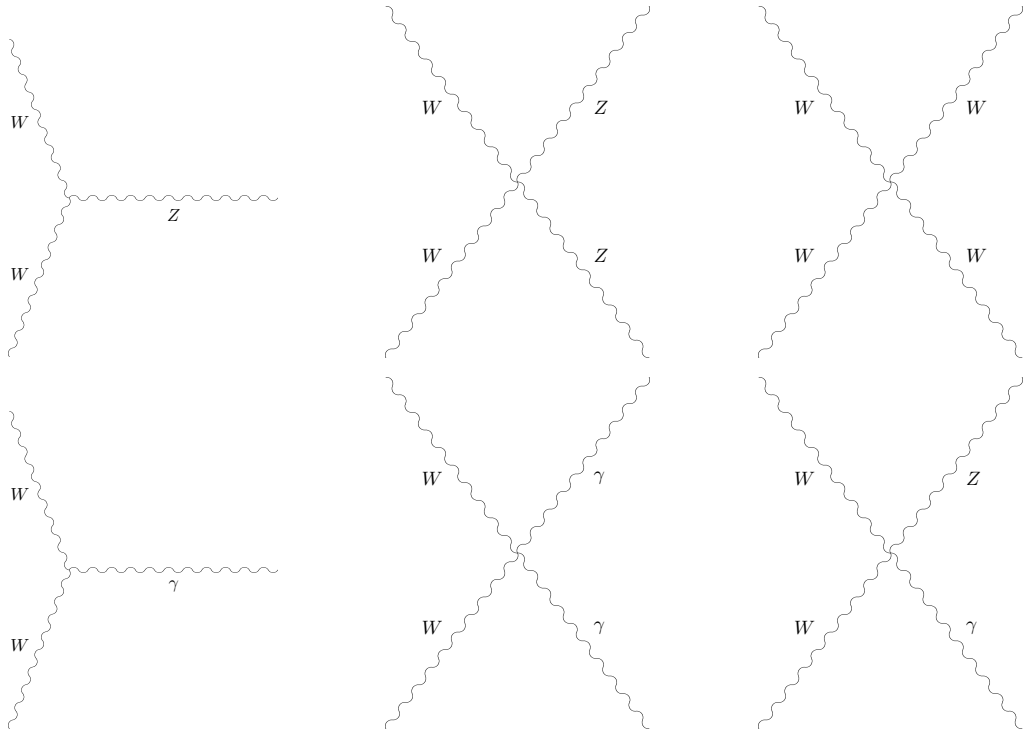

Figure 6: Interaction vertices involving only intermediate vector bosons and photons.

quarks but only within a family, and it would therefore be possible to define a conserved family quark 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. In other words, the quark mass eigenstates do not coincide with the 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 a down (anti)quark. The unitary matrix that determines how flavours mix is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

Although there is no exactly conserved quark family number, there is an approximately conserved quark flavour number, which does not change as long as weak interactions are not involved in the process: flavour is in fact exactly conserved by strong and electromagnetic interactions. When weak interactions are also considered this is not true anymore due to flavour

quantity conserved\interaction	EM	strong	weak
lepton type	yes	–	no
flavour	yes	yes	no
lepton family	yes	–	yes (if massless)
quark family	yes	yes	no
lepton number	yes	–	yes
quark number	yes	yes	yes

Table 5: Conservation laws - part 1.

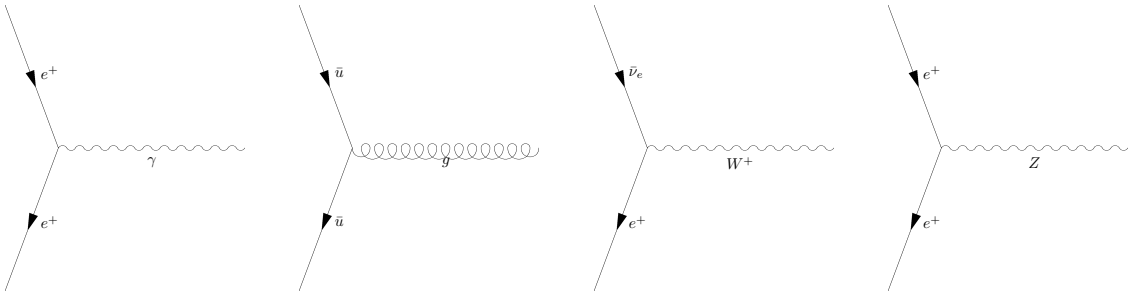

Figure 7: The fundamental vertices involving matter antiparticles.

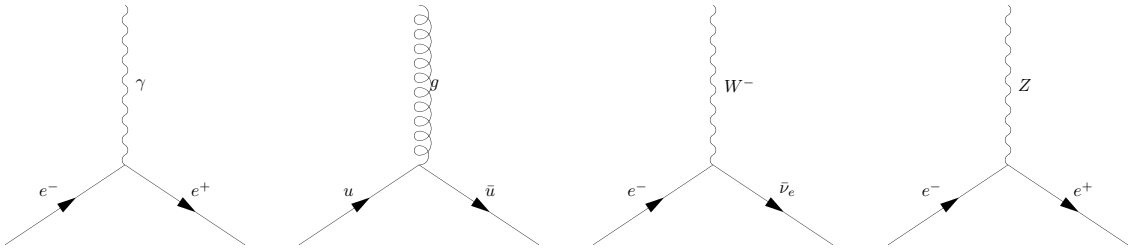

Figure 8: The fundamental vertices involving both particles and antiparticles.

mixing, but the total number of quarks minus that of antiquarks, $\mathcal{Q}$, is still exactly conserved. This is usually express in terms of conservation of the baryon number $\mathcal{B}$, i.e. the number of baryons minus the number of antibaryons. Indeed, the number of baryons is just three times the number of quarks, and the number of antibaryons three times the number of antiquarks. Each meson contributes zero to the total quark number, as it involves a quark and an antiquark. Therefore, $\mathcal{Q} = 3\mathcal{B}$, and the two conservation laws are equivalent. Baryon number is usually traded for one of the conserved flavour quantum numbers, while another one is traded for electric charge. Denoting with f and $\bar{f}$ the number of quarks and antiquarks of flavour f, respectively, we clearly have $\mathcal{B} = \sum(f - \bar{f})/3$, while for the electric charge $Q = (2/3)(u - \bar{u} + c - \bar{c} + t - \bar{t}) - (1/3)(d - \bar{d} + s - \bar{s} + b - \bar{b})$. Strangeness is defined, for historical reasons, as $S = \bar{s} - s$, while for charm and beauty one has more logically $C = c - \bar{c}$ and $B = b - \bar{b}$. A summary of the conservation laws related to particle types is given in Table 5. 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.

The diagrams of Figs. 4 to 8 describe how the interaction works at the most fundamental level, but cannot represent a true physical process due to energy-momentum conservation. To describe an actual physical process we have to properly combine two or more of them. Let us begin with the electromagnetic interactions. If we combine two electromagnetic vertices together like in the left panel of Fig. 9, then we are representing the scattering process of two electrons (Møller scattering). The second panel correspond to the same process: 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. In this case there is no problem with the emission/absorption process: while energy and momentum are conserved, the exchanged photon is not on its *mass shell*, i.e., $p_\gamma^2 \neq 0$. Internal lines in Feynman diagrams represent

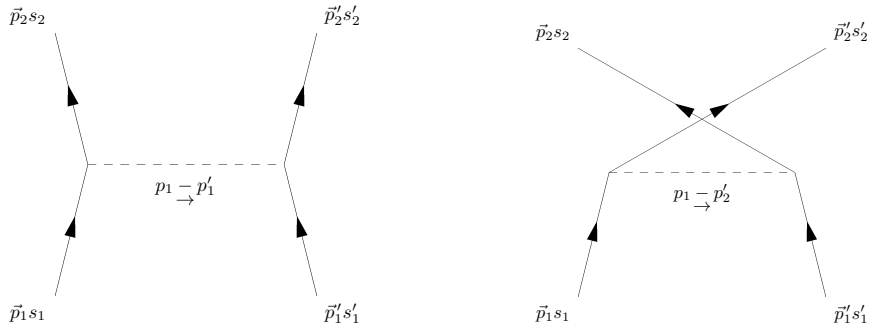

Figure 9: Møller scattering.

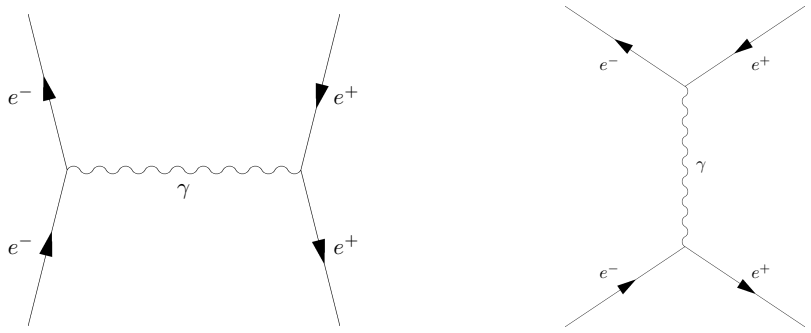

Figure 10: Bhabha scattering.

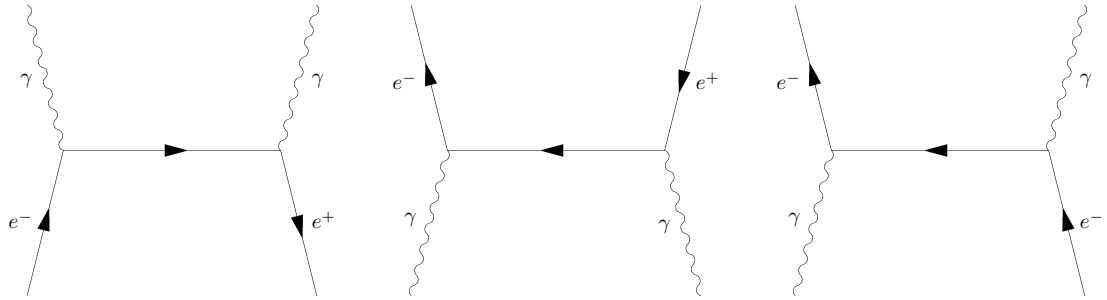

Figure 11: Left: e^-- e^+ annihilation. Center: pair production. Right: Compton scattering.

virtual particles, not real particles, which do not have to be on shell. External lines, on the other hand, represent the real particles that initiate or come out of a physical process.

If we now reflect the right fermionic line in the first panel, we obtain a diagram describing electron-positron scattering (Bhabha scattering) via the exchange of a photon, see the left panel of Fig. 10. This kind of process can take place also through a different route: the electron and the positron can annihilate into a photon, and a new pair will be created from the photon (right panel of Fig. 10).

There are three more ways to combine two QED vertices. These are shown in Fig. 11, and correspond (from left to right) to electron-positron annihilation, electron-positron pair creation, and Compton scattering.

It is important at this point to make it clear that Feynman diagrams provide only a schematic

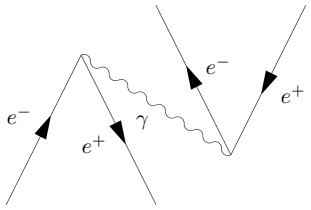

Figure 12: The same diagram as in the right panel of Fig. 10, only drawn in a different way.

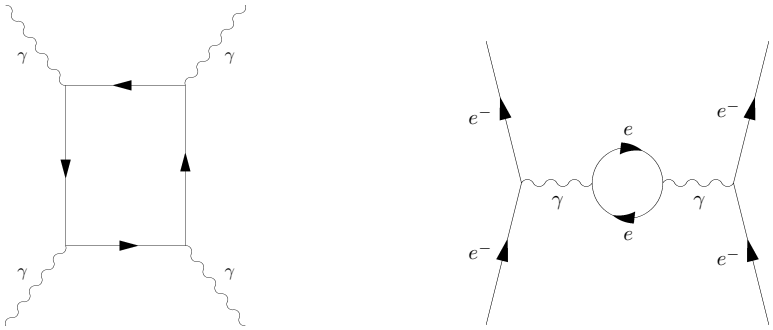

Figure 13: More complicated diagrams.

description of the process from the visual point of view, *not* an accurate depiction of particle trajectories. The notion of time that we are using is also somewhat misleading. Consider the diagram of Fig. 12: does it represent a different process than the one in the right panel of Fig. 10? 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 exacly the annihilation of the first pair happens, and when the creation of the second pair happens, are actually meaningless questions.

There is of course an infinity of ways in which one can combine an indefinite number of vertices. A couple of examples is shown in Fig. 13. The left panel shows the light-by-light (Delbrück) scattering process, while the right panel corresponds again to electron-electron scattering, with the exchanged photon turning temporarily into an electron positron pair. As we will learn later, diagrams have different "weights", i.e., are more or less important in the description of a process. Each vertex contributes a factor e to the weight of a diagram, so the diagrams of Figs. 9, 10 and 11 are all proportional to $e^2 = 4\pi\alpha$ (in natural units), with $\alpha \sim 1/137$ the *fine structure constant*. The diagrams of Fig. 13 are proportional to α^2, and so are relatively suppressed. This means that to describe a given process to a given precision we will need only a limited number of diagrams. The factor weighting each vertex is called *coupling constant*: there is an electromagnetic one (i.e., the absolute value of the electron charge), a strong one, and two weak ones (one for the charged process and one for the neutral process), which are however related to the electromagnetic coupling through two functions of a single parameter, known as *Weinberg angle*, as a consequence of electroweak unification.

Let us consider now processes mediated by the weak interactions. The simplest such process is the decay of a muon (Fig. 14, left): $\mu^- \to e^- \nu_\mu \bar{\nu}_e$ (notice the conservation of both the electronic and the muonic lepton numbers). Except for the presence of spectator quarks, the

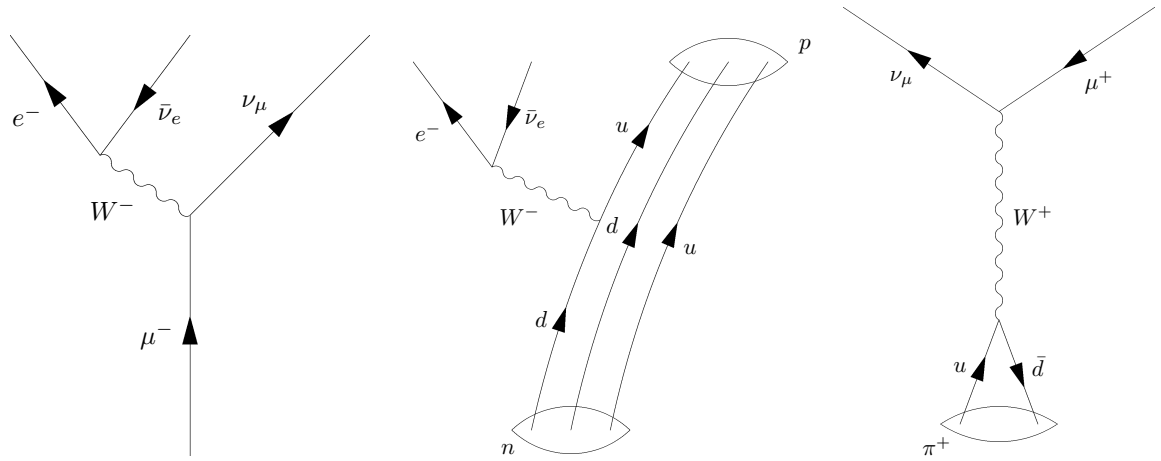

Figure 14: Left: muon decay. Center: Beta decay of the neutron. Right: Charged pion decay.

same diagram describes the beta decay of the neutron (Fig. 14, center). The same diagram with electron and electron neutrino replaced by muon and muonic neutrino describes the main decay mode of the charged pion (Fig. 14, right). Reflecting vertically the diagram for the β decay of the neutron we obtain the diagram for the process of antineutrino capture, $\bar{\nu}_e + p \to e^+ + n$. The decays of the neutron and of the pion represent leptonic decays of hadrons, but weak interactions are responsible also for non-leptonic decay modes. For example, the Λ^0 (uds) decays to $p + \pi^-$ via the emission of a W^- from the s quark, which turns into a u; the quarks produced in the decay $W^- \to \bar{u}d$ correspond precisely to the content of π^-, while the remaining quarks can form a proton. In this case the *strangeness* S of the baryon (which is the number of strange antiquarks minus the number of strange quarks) changes by $\Delta S = 1$. Hadrons can also decay via electromagnetic interactions, for example $\Sigma^0(uds) \to \Lambda^0(uds)\gamma$, which has $\Delta S = 0$ (but $\Delta I = -1$, where I is the isospin: more on this later). Weak interactions contribute also to processes where the flavour content of a baryon does not change. For example, the Δ^+ (uud) can decay into $n + \pi^+$ via the emission of a W^+ from one of the u quarks, which then turns into a d quark, and subsequent decay of the W^+ into $u + \bar{d}$: the $\bar{d}$ combines with a u to form a π^+, while the remaining quarks form a neutron. However, the same decay can take place via strong interactions, with a $d\bar{d}$ pair created by a gluon emitted by one of the initial quarks. This process has actually a much larger weight, and so it is the one actually governing the decay process.

To conclude this subsection, it is worth discussing an important detail concerning strong interactions. Although the fundamental interaction vertices involve quarks and gluons, the phenomenon of confinement implies that the correct effective description of strong interactions in terms of exchange of particles should be done using hadrons. The lightest mediator is therefore not the massless gluon, but the massive (although relatively light) pion.

1.8 Range and strength of the interactions

Now that we know (at least qualitatively) how interactions work, we can estimate their range and strength theoretically. We begin by discussing the range, i.e., "how far they are felt" in space. The basic idea is that the range of interaction is the inverse of the mass of the lightest mediator. This can be understood at a very qualitative level using the uncertainty relation

for energy, $\Delta E \Delta t \geq \frac{\hbar}{2}$: the exchange of a particle of mass M requires a "violation" of energy conservation over a time $\Delta t \sim \hbar/\Delta E \sim \hbar/(Mc^2)$, over which it can travel a distance not larger than $\Delta x = c\Delta t$ (if it moved at the speed of light). All in all, $\Delta x \sim \hbar/(Mc)$, which is the Compton length of the mediator. This argument is clearly a back-of-an-envelope calculation for several reasons, so let us try to make it more quantitative.

We begin by noticing that in the nonrelativistic, free-particle case the fundamental dynamical equation, i.e., the Schrödinger equation, can be derived from the relation between energy and momentum, $E = \frac{\vec{p}^2}{2m}$, by upgrading the classical variables to operators acting on the particle's wave function via the replacements $E \to i\partial_t$ and $\vec{p} \to -i\vec{\nabla}$,

$$E = \frac{\vec{p}^2}{2m} \Rightarrow i\partial_t \psi(t, \vec{x}) = -\frac{\vec{\nabla}^2}{2m}\psi(t, \vec{x})\,. \tag{1.7}$$

Energy and momenta are replaced by time and space derivatives, i.e., by the generators of temporal and spatial translations. In the relativistic case the energy-momentum relation is instead $E^2 - \vec{p}^2 = M^2$ which leads, after the same substitutions for E and $\vec{p}$, to the *Klein-Gordon equation*, which is essentially the relativistic version of the Schrödinger equation:

$$(\Box + M^2)\psi(t, \vec{x}) = (\partial_t^2 - \vec{\nabla}^2 + M^2)\psi(t, \vec{x}) = 0\,. \tag{1.8}$$

Unfortunately, for a correct description of relativistic particles it is not enough to simply make relativistic the energy-momentum relation but, as we will see in due course, it is also necessary to change how particles are described. In fact, the nonrelativistic wavefunction ψ has to be replaced with a quantum field $\hat{\Psi}$, i.e., a field *operator* acting on the Hilbert space of the particle. A quantum field obeying the Klein-Gordon equation describes the propagation of free particles of mass M. Perhaps the simplest way to get accostumed to this is to recall that as an equation for a classical field in the case $M = 0$, Eq. (1.8) is simply Maxwell's equation for the electromagnetic field in the vacuum, i.e., $\Box A_\mu = 0$. Upon quantisation, the electromagnetic field describes massless quanta - the photons. For nonzero M and upon quantisation one would therefore expect this equation to describe the propagation of particles of mass M - and that is precisely what it does.

If we now consider the massive case and follow the reverse route, i.e., we go back to the classical case, Eq. (1.8) will describe the propagation of some classical field mediating some interaction, in the same way as the photon field does with the electromagnetic one. Therefore, solving the Klein-Gordon equation as an equation for a *classical* (relativistic) field will tell us what the interaction mediated by our massive particle looks like when the classical limit is approached, which corresponds to the case of the exchange of a large number of particles. In particular, we can determine the classical potential corresponding to the massive particle by solving the Klein-Gordon equation in the static (i.e., time-independent) case in the presence of a fixed source (much in the same way as we obtain the Coulomb potential putting a static electric source somewhere in space):

$$(-\vec{\nabla}^2 + M^2)u(\vec{x}) = g\delta(\vec{x})\,, \tag{1.9}$$

where g is the "charge" of the source. This equation is most easily solved after a Fourier transformation to momentum space,

$$u(\vec{x}) = \int \frac{d^3p}{(2\pi)^3}\, e^{i\vec{p}\cdot\vec{x}}\tilde{u}(\vec{p})\,. \tag{1.10}$$

In momentum space we find

$$(\vec{p}^2 + M^2)\tilde{u}(\vec{p}) = g \Rightarrow \tilde{u}(\vec{p}) = \frac{g}{\vec{p}^2 + M^2}\,. \tag{1.11}$$

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

$$\begin{aligned} u(\vec{x}) &= \int \frac{d^3p}{(2\pi)^3}\, e^{i\vec{p}\cdot\vec{x}} \frac{g}{\vec{p}^2 + M^2} = \frac{g}{(2\pi)^2} \int_0^\infty dp\, p^2 \int_{-1}^{+1} dz\, e^{iprz} \frac{1}{p^2 + M^2} \\ &= \frac{g}{(2\pi)^2 ir} \int_0^\infty dp\, p\, (e^{ipr} - e^{-ipr}) \frac{1}{p^2 + M^2} = \frac{g}{(2\pi)^2 ir} \int_{-\infty}^\infty dp\, p\, e^{ipr} \frac{1}{p^2 + M^2} \\ &= \frac{g}{(2\pi)^2 ir} (2\pi i) e^{-Mr} \frac{iM}{2iM} = \frac{g}{4\pi r}\, e^{-Mr}\,, \end{aligned} \tag{1.12}$$

where we have made use of the residue theorem to compute the last integral. The potential $u(r) = \frac{g}{4\pi r}\, e^{-Mr}$ is the *Yukawa potential*, and it clearly has range $1/M$. Putting back the appropriate factors of $\hbar$ and c we get

$$\text{range} = \frac{\hbar}{Mc} = \frac{\hbar c}{Mc^2} = \frac{197\ \text{MeV} \cdot \text{fm}}{M[\text{MeV}/c^2]\text{MeV}} = \frac{197}{M[\text{MeV}/c^2]}\,\text{fm}\,. \tag{1.13}$$

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

$$\text{range}_{\text{weak}} = \frac{197}{8 \cdot 10^4}\,\text{fm} = 2.5 \cdot 10^{-3}\,\text{fm}\,. \tag{1.14}$$

As we have remarked above, due to confinement the actual mediator of the strong force is the pion, and so (using $M_{\pi^0} = 135\,\text{MeV}/c^2$)

$$\text{range}_{\text{strong}} = \frac{197}{135}\,\text{fm} = 1.5\,\text{fm}\,, \tag{1.15}$$

which is actually the typical scale of nuclei. The inverse of the range of interaction gives an estimate of the typical energy scale of the strong and weak interactions, respectively $\mathcal{O}(100\text{MeV})$ and $\mathcal{O}(100\text{GeV})$. Electromagnetic interactions are mediated by massless photons, which entails an infinite range: indeed, in this limit the Yukawa potential reduces to the Coulomb potential $\sim 1/r$.

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 particles decaying through different interactions. As a matter of fact, Feynman diagrams provide more than a schematic description of physical processes. We have already said that each diagram has an associated weight: such a weight, which is technically called *amplitude*, allows to compute the probability of the corresponding process, roughly via the relation decay width/scattering rate (probability per unit time) $\propto |\text{amplitude}|^2$. In the case of a decay process we then obtain the decay width $\Gamma = \tau^{-1}$. The simplest diagrams for decay processes typically involve two vertices, and so Γ is typically proportional to the fourth power of the coupling constant (the weight factor coming with each vertex), and so

$$\frac{g_1^4}{g_2^4} \sim \frac{\tau_2}{\tau_1} \Rightarrow \frac{g_1^2}{g_2^2} \sim \sqrt{\frac{\tau_2}{\tau_1}}\,. \tag{1.16}$$

Typical decay times for strong interactions are in the range $10^{-23} \div 10^{-20} s$, those for weak interactions are in the range $10^{-13} \div 10^{3} s$, while for electromagnetic interactions they are $10^{-16} s$. To compare strong and weak interactions we can for example compare the decay of the Δ^0, which takes place through strong interactions with $\tau_\Delta = 5.6 \cdot 10^{-24} s$, to that of the neutron, which takes place through weak interactions with $\tau_n = 880\, s$:

$$\frac{g_w^2}{g_s^2} \sim \sqrt{\frac{\tau_\Delta}{\tau_n}} \sim 10^{-13}\,. \tag{1.17}$$

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

$$\frac{g_w^2}{g_{em}^2} \sim \sqrt{\frac{\tau_{\pi^0}}{\tau_{\pi^+}}} \sim 10^{-4} \div 10^{-3}\,. \tag{1.18}$$

Different estimates will be obtained comparing different processes, but the bottom line is that the strong force is stronger than the electromagnetic one which is stronger than the weak one.

A similar kind of estimate can be done looking at scattering processes. In that case, the cross section of a process is proportional to the absolute value squared of the amplitude associated to Feynman diagrams, and therefore, to lowest order, to the fourth power of the coupling constant. We then find that the stronger the interaction, the larger the typical cross section for a processes mediated by it.

It is worth pointing out 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. In fact, in the Glashow-Weinberg-Salam theory that describes weak and electromagnetic interactions, 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 will learn later, the internal boson lines in the Feynman diagrams correspond to factors $(p^2 - m^2)^{-1}$, where p and m are respectively the four-momentum and the mass of the boson. Effectively, the exchange of a W boson between a pair of electrons with energies much lower than the W mass comes with a factor $G_F = g_w^2/M_W^2 \sim 10^{-5}$ GeV^{-2}: this is the coupling constant actually estimated above. Weak interactions are therefore weak not because of a small coupling, but because of the large mass of the mediators. 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.

A historical note: the quantity G_F is known as the *Fermi constant*, and it appeared in the first theory of β decays proposed by Fermi in 1933. In that theory the proton, neutron, electron and neutrino interacted via a four-fermion vertex, with coupling constant G_F. From the modern perspective, the four-fermion vertex can be considered as an effective description valid at energies much lower than M_W. The fact that G_F has dimensions of some inverse power of mass indicated 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). The reason for this will be discussed much later. 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 $M_W \sim \sqrt{4\pi\alpha/G_F} \sim 30\,\text{GeV}$, which is only 2.5 times smaller than the actual value.

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

We have discussed above how diagrams built using vertices from the various interaction can describe different physical processes. Is there a way to tell what kind of interaction will be responsible for a given decay process even before drawing the relevant diagrams? The answer is certainly yes: 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. If we see pions there are good chances that we are dealing with a strong process, but we cannot be sure, as will become clear below. However, not always there is a clear signature, and in those cases we have to rely on other clues. We have already seen one of these, namely the lifetime of a decaying particle. There is another thing we can look at: the symmetry properties of a particular process.

As we have already mentioned, in general, unless prevented by some conservation law, particles decay into lighter particles. Certain conservation laws are valid for all interactions, namely conservation of energy, momentum, and angular momentum, as they are just consequences of translation and rotation invariance. 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. These conservation laws imply kinematical constraints on the decay processes. Other conservation laws are instead of dynamical nature, deriving from the detailed nature of the interactions discussed in the previous section. We have seen there that all interactions conserve electric charge, baryon number and lepton number. We have also seen that lepton family numbers are approximately conserved by all interactions. Other conservation laws are valid for certain interactions and not for others, and this allows to distinguish what is the interaction responsible for a given process. In particular, conservation laws related to the type of particles have been listed in Table 5. We want to remark that although we began the discussion with decay processes, of course conservation laws do apply to scattering processes as well.

We want to give now a brief overview of the symmetries of the various interactions starting from the strong ones, which are the "most symmetrical" of the interactions. As we have already remarked, strong interactions conserve *flavour*, i.e., the number of quarks minus the number of antiquarks of a certain type is conserved in a process. The number of u and d quarks can be traded for baryon number and electric charge, which we have already said are conserved by all interactions. The *strangeness* S (number of strange antiquarks minus number of strange quarks - the choice of signs looks strange, indeed, but that is how it came about historically), *charm* (number of charm quarks minus charm antiquarks), and so on are conserved also by electromagnetic interactions, but *not* by the weak interactions, which allow flavour to change. For this reason, pions can appear in the final state due to a flavour-changing weak current: for example in the decay of neutral kaons: in that case, it is the violation of strangeness (as well as the long lifetime) that tells us that we are not dealing with strong interactions.

Strong interactions also conserve approximately a quantity called *isospin*. For reasons that we will discuss later, one can assign an isospin I to multiplets of light hadrons (i.e., those made of u, d and s), within which they are distinguished by a second number I_3. All other particles are assigned $I = 0$. Does the thing look somehow familiar? As a matter of fact isospin shares a lot of properties with spin, although it is entirely unrelated to it. For example, the three pions form a triplet to which it is assigned the value $I = 1$, with I_3 equal to their charge. 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

quantity	strong	em	weak
Q	y	y	y
$\mathcal{B}$	y	y	y
L	y	y	y
I	y	n	n
S	y	y	n
P	y	y	n
C	y	y	n
T/CP	y	y	n
CPT	y	y	y

Table 6: Conservation laws - part 2.

with $I = \frac{3}{2}$. 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 scattering processes. Notice that both electromagnetic and weak interactions do *not* conserve this quantity (pion decay modes should convince you of that).

There are three important discrete symmetries that we have to discuss, namely parity P (i.e., spatial inversion), charge conjugation C (exchange of particle with antiparticle) and time reversal T. As it turns out, strong and electromagnetic interactions conserve all three of them, while weak interactions do not conserve any. At the same time, the combination CPT is conserved by all the interactions: this is actually a theorem in quantum field theory, so that conservation of T is equivalent to conservation of CP. The situation is summarised in Table 6. 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 everything, then it is taking place through the strong interactions.

It is worth spending a few words on the meaning of parity non-conservation. Weak interactions are not invariant under parity, which means that the physical phenomena that we see in the mirror are not always possible physical phenomena in our, real, world. The strongest evidence of parity non-conservation is the fact that all neutrinos are left-handed and all antineutrinos right-handed, which means in practice that they have helicity -1 and $+1$, respectively. Helicity is the spin component in the directions of motion, $h = \frac{\vec{s}\cdot\vec{p}}{|\vec{p}|}$, which for massless particles is a Lorentz-invariant quantity. Under parity helicity changes sign, so that in the mirror we would see a right-handed neutrino, which does not exist (or at least has never been observed) in our world. Right after the suggestion of Lee and Yang that parity might not be conserved in weak interactions (1956), the first observation of parity non-conservation was in the decay of cobalt-60, $^{60}\mathrm{Co} \to {}^{60}\mathrm{Ni} + e^- + \bar{\nu}_e$. Here the electron is emitted preferentially in the direction opposite to the spin of the initial nucleus (Wu *et al.*, 1956). The picture provided by a mirror parallel to the nuclear spin axis is obtained by combining parity and a 180° rotation around an axis orthogonal to the mirror: the spin is reversed, and electrons look like being preferentially emitted in the direction of the spin. Shortly thereafter, it was observed that in the decay of polarised muons, the electron distribution is proportional to $1 - \frac{1}{3}\cos\theta$, with θ the angle between the electron direction and the muon polarisation (Garwin *et al.*, 1956). Under parity the spin of the muon does not change, but the momentum of the electron is reversed, and thus it become angularly distributed according to a $1 + \frac{1}{3}\cos\theta$ law. By the way, this is the law governing the angular distribution of the positron in the decay of the antimuon, and this shows that also charge con-

jugation invariance is violated in muon decay. Combined parity and charge conjugation seems still a good symmetry, but as it turns out it is also violated by weak interactions.

There is another quantity that we have seen it is conserved at interaction vertices, and that is colour. What are the visible consequences of this conservation law? Actually none: confinement of quarks and antiquarks in hadrons come with the property that hadrons have net colour zero, and so the conservation law boils down to zero colour in, zero colour out. 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 has been found to be of the form

$$V_s = -\frac{4}{3}\frac{\alpha_s}{r} + \sigma r\,, \tag{1.19}$$

in the limit of infinite quark masses, where $\alpha_s = \frac{g_s^2}{4\pi}$ with g_s the strong coupling constant, and σ the *string tension*. As the quark and the antiquark are pulled apart, the energy stored in the system keeps increasing, 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, at some point the energy stored in the system is sufficient to create a $q\bar{q}$ pair out of the vacuum, with the new particles binding to the old ones. This process is called *string breaking*.

2 Symmetries

In this section we discuss in some detail the concept of symmetry in a physical theory and its consequences.

Quoting almost verbatim from Weinberg's book [2], a *symmetry* is a change in the experimenter's point of view that does not change the results of possible experiments. Let us explain this in more detail. Consider two experimenters $\mathcal{O}$ and $\mathcal{O}'$ making measurements on the same physical system. They subscribe to the same operative rules concerning the measurement of the various observables, but they use in general different reference frames, so that in general they find different values for the various physical quantities, thus producing two different descriptions of the same system. In mathematical terms, this means that they will assign to the system different representative vectors in the corresponding Hilbert space. This is because the expectation values of the operators corresponding to the physical observables (which are the *same* operators for both observers) are different, reflecting the different results they have obtained. More precisely, $\mathcal{O}$ and $\mathcal{O}'$ will assign respectively the *rays*[11] $\mathcal{R}$ and $\mathcal{R}'$ to the state of the system. Although the two descriptions are in general not the same, for certain pairs of observers they will be *equivalent*, i.e., the physical laws implied by the measurements will be the same for both observers. In other words, it will be impossible for an observer to determine her or his reference frame using only her or his measurements. If two observers give equivalent descriptions, then the set of possible physical states that they can observe must be the same. Mathematically, the set of rays that they can assign to the system must be the same. Moreover, if $\mathcal{O}$ sees two states of the system as different, so must do $\mathcal{O}'$.

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

[11] A ray is an equivalence class of vectors $\psi \in \mathcal{H}$ with respect to the equivalence relation $\{\psi \sim \phi \text{ if } \psi = e^{i\zeta}\phi\}$.

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. We then have a mapping $\mathcal{M} : \underline{\mathcal{H}} \to \underline{\mathcal{H}}$ 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 they would not be equivalent.

Suppose that we now perform an experiment on the system, and we see it transition from a state to another. The two observers will see the transitions

$$\mathcal{O} : \mathcal{R}_i \longrightarrow \mathcal{R}_f \,, \qquad \mathcal{O}' : \mathcal{R}'_i \longrightarrow \mathcal{R}'_f \,, \tag{2.1}$$

occurring with probabilities P and P',

$$P = (\mathcal{R}_i \cdot \mathcal{R}_f)^2 \,, \qquad P' = (\mathcal{R}'_i \cdot \mathcal{R}'_f)^2 \,, \tag{2.2}$$

where

$$\mathcal{R}_1 \cdot \mathcal{R}_2 = |(\psi_1, \psi_2)| \,, \tag{2.3}$$

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 process, the transition probabilities they observe must be the same, $P = P'$, and so

$$\mathcal{R}_i \cdot \mathcal{R}_f = \mathcal{R}'_i \cdot \mathcal{R}'_f \,. \tag{2.4}$$

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

$$\mathcal{R}_i \cdot \mathcal{R}_f = (\mathcal{M}\mathcal{R}_i) \cdot (\mathcal{M}\mathcal{R}_f) \,. \tag{2.5}$$

A theorem due to Wigner guarantees that an invertible transformation $\mathcal{M}$ on the space of rays $\underline{\mathcal{H}}$ that conserves probabilities can be implemented as a transformation on the space of vectors $\mathcal{H}$ that is either linear and unitary or antilinear and antiunitary:

$$\begin{aligned} \text{linear and unitary :} \quad & U(\alpha\psi + \beta\phi) = \alpha U\psi + \beta U\phi \,, \qquad (U\psi, U\phi) = (\psi, \phi) \,, \\ \text{antilinear and antiunitary :} \quad & T(\alpha\psi + \beta\phi) = \alpha^* T\psi + \beta^* T\phi \,, \qquad (T\psi, T\phi) = (\psi, \phi)^* \,. \end{aligned} \tag{2.6}$$

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. For a proof of the theorem, confer [2].

Here is a brief detour on unitary and antiunitary operators. Unitary and antiunitary transformations are by definition norm-preserving and onto transformations, meaning that they *i)* preserve the norm $\|\psi\|$ of vectors and *ii)* have the whole Hilbert space as their image: they only differ in being linear or antilinear. Property *i)* implies $U^\dagger U = T^\dagger T = \mathbf{1}$: this follows from the rightmost column of Eq. (2.6), which in turn is a consequence of norm preservation and of the polarisation identity,

$$4(\psi, \phi) = \|\psi + \phi\| + \|\psi - \phi\| + i\|\psi - i\phi\| - i\|\psi + i\phi\| \,. \tag{2.7}$$

Here the dagger denotes the adjoint of an operator, defined in the linear and antilinear case respectively as

$$(\psi, U\phi) = (U^\dagger\psi, \phi) \,, \qquad (\psi, T\phi) = (T^\dagger\psi, \phi)^* \,. \tag{2.8}$$

Property *ii)* implies that $UU^\dagger = TT^\dagger = \mathbf{1}$: indeed, since for every ψ there is a ϕ such that $\psi = U\phi$, then $\psi = U(U^\dagger U)\phi = (UU^\dagger)U\phi = (UU^\dagger)\psi$. The same argument applies to T. Conversely, if $U^\dagger U = UU^\dagger = \mathbf{1}$ and $T^\dagger T = TT^\dagger = \mathbf{1}$ for linear U and antilinear T, then it follows norm preservation (from the first identity), and the trivial identity $\psi = UU^\dagger\psi = U(U^\dagger\psi)$ implies that these transformations are onto.

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.

Suppose for generality that our symmetry transformation be time-dependent, being represented by a time-dependent operator $M(t)$. What we request is that

$$M(t)U(t)\psi(0) = U(t)M(0)\psi(0)\,, \tag{2.9}$$

where the temporal evolution of the system is given by the unitary operator $U(t) = e^{-iHt}$ with H a time-independent Hamiltonian, and $\psi(0)$ is the state vector at $t = 0$ for the observer $\mathcal{O}$. Eq. (2.9) expresses precisely 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

$$M(t)U(t) = U(t)M(0) \Rightarrow M(t) = U(t)M(0)U(t)^\dagger\,. \tag{2.10}$$

Why does it have to be so? In general we can write

$$M(t)U(t)\psi(0) = M(t)U(t)M(0)^\dagger M(0)\psi(0) = \mathcal{U}(t)M(0)\psi(0)\,, \tag{2.11}$$

where $\mathcal{U}(t)$ provides the unitary temporal evolution of the system as seen by the observer $\mathcal{O}'$. If $\mathcal{U}(t)$ is the same operator as $U(t)$, then the temporal evolution of the system obeys the same laws for both observers, and all that change is the initial condition, i.e., the state vector at $t = 0$, which is different for the two observers because the results of their measurements are. If we were to do an experiment as observer $\mathcal{O}$ starting with $\psi(0)$, and another experiment as observer $\mathcal{O}'$ starting with $\psi(0)$ (which means that $\mathcal{O}$ would use the vector $M(0)^\dagger\psi(0)$ instead), then the results obtained would be the same for the two observers in the two (different!) experiments. In general, however, $\mathcal{U}(t)$ differs from $U(t)$. We can obtain $\mathcal{U}(t)$ by writing down the differential equation it obeys, and imposing the initial condition $\mathcal{U}(0) = \mathbf{1}$. We have

$$\begin{aligned}\frac{d\mathcal{U}(t)}{dt} &= \frac{dM(t)}{dt}U(t)M(0)^\dagger + M(t)(-iHU(t))M(0)^\dagger \\ &= -i\left(M(t)HM(t)^\dagger + i\frac{dM(t)}{dt}M(t)^\dagger\right)\mathcal{U}(t) \equiv -i\mathcal{H}_M(t)\mathcal{U}(t)\,,\end{aligned} \tag{2.12}$$

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 follows from the fact that $M(t)$ is norm-preserving and onto:

$$0 = \frac{d}{dt}\mathbf{1} = \frac{d}{dt}[M(t)M(t)^\dagger] = \frac{d}{dt}[M(t)]M(t)^\dagger + M(t)\frac{d}{dt}[M(t)^\dagger\,, \tag{2.13}$$

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. (2.12) is the time-ordered exponential of $\mathcal{H}_M(t)$,

$$\begin{aligned}\mathcal{U}(t) &= \mathrm{Texp}\left\{-i\int_0^t dt'\,\mathcal{H}_M(t')\right\}\\ &\equiv \sum_{n=0}^{\infty}(-i)^n\int_0^t dt_1'\int_0^{t_1'} dt_2'\ldots\int_0^{t_{n-1}'} dt_n'\mathcal{H}_M(t_1')\mathcal{H}_M(t_2')\ldots\mathcal{H}_M(t_n')\\ &= \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_0^t dt_1'\int_0^t dt_2'\ldots\int_0^t dt_n'\mathrm{T}\left(\mathcal{H}_M(t_1')\mathcal{H}_M(t_2')\ldots\mathcal{H}_M(t_n')\right)\,,\end{aligned} \tag{2.14}$$

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

$$\mathrm{T}\left(A(t_1)\ldots A(t_n)\right) = \theta(t_1-t_2)\ldots\theta(t_{n-1}-t_n)A(t_1)\ldots A(t_n) + \text{permutations}\,. \tag{2.15}$$

Clearly, if $\mathcal{H}_M(t) = H$ then $\mathcal{U}(t) = U(t)$ (they solve the same 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

$$H = M(t)HM(t)^\dagger + i\frac{dM(t)}{dt}M(t)^\dagger \Rightarrow -i[H, M(t)] = \frac{dM(t)}{dt}\,, \tag{2.16}$$

and this is solved by

$$M(t) = e^{-iHt}M(0)e^{iHt} = U(t)M(0)U(t)^\dagger \Rightarrow M(t)U(t) = U(t)M(0)\,. \tag{2.17}$$

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. (2.17) 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.

There is an important structure underlying the set of symmetry transformations of a physical system. If M_1 and M_2 are (possibly time dependent) symmetry transformations, so is their composition $M = M_2M_1$. Indeed, the product of unitary and/or antiunitary operators is still either unitary or antiunitary, and moreover

$$M(t) = M_2(t)M_1(t) = U(t)M_2(0)M_1(0)U(t)^\dagger = U(t)M(0)U(t)^\dagger\,. \tag{2.18}$$

Such a composition is associative, i.e., if M_3 is another symmetry transformation then $M_3M_2M_1 = M_3(M_2M_1) = (M_3M_2)M_1$. 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.

In general, symmetries are classified in two big families, namely *continuous* and *discrete* symmetries: in the first case there is a continuous family of symmetry transformations depending on some real parameter, in the second case the transformation is "isolated". If $M = M(\alpha)$ is an element of a continuous family of symmetry transformations dependent on some parameter

α and connected to the identity, $M(0) = \mathbf{1}$, then it has to be unitary. Indeed, since $f(\alpha) = (M(\alpha)\psi, M(\alpha)\phi)$ has to be equal to (ψ, ϕ) or $(\psi, \phi)^*$, it must be a constant and so $f(\alpha) = f(0) = (\psi, \phi)$.

Let us now make our life simpler, and let us focus on time-independent symmetry transformations. From Eq. (2.17) we find $[U(t), M] = 0$ at all times, and so taking the time derivative we find

$$[H, M] = 0\,. \tag{2.19}$$

A time-independent transformation is therefore a symmetry if it commutes with the Hamiltonian. Let us focus on a continuous family of symmetry transformations $M(\alpha)$ forming a one-parameter group,[12]

$$M(\alpha_1)M(\alpha_2) = M(\alpha_1 + \alpha_2)\,, \tag{2.20}$$

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

$$\begin{aligned} M(d\alpha)M(\alpha) &= M(\alpha + d\alpha)\,, \\ \left(M(0) + d\alpha \frac{dM}{d\alpha}(0)\right) M(\alpha) &= M(\alpha) + d\alpha \frac{dM}{d\alpha}(\alpha)\,, \\ \frac{dM}{d\alpha}(0)M(\alpha) &= \frac{dM}{d\alpha}(\alpha)\,. \end{aligned} \tag{2.21}$$

This differential equation is easily solve to give

$$M(\alpha) = \exp\left\{\alpha \frac{dM}{d\alpha}(0)\right\} = \exp\left\{i\alpha(-i)\frac{dM}{d\alpha}(0)\right\} = \exp\left\{i\alpha Q\right\}\,. \tag{2.22}$$

Since $M(\alpha)$ is unitary, it follows that $Q = Q^\dagger$ is self-adjoint. Since $M(\alpha)$ is a symmetry for all α, we can take the derivative of $0 = [H, M(\alpha)]$ with respect to α at $\alpha = 0$ to find $[H, Q] = 0$. This means that Q is a conserved physical quantity that can be diagonalised simultaneously with the Hamiltonian. To better appreciate the meaning of what a conserved quantity is, it is useful to switch from the *Schrödinger picture* used so far, in which observables are time-independent and state vectors depend on time, to the *Heisenberg picture* in which state vectors are time-independent and observabels depend on time. This is done by rewriting the expectation value of an observable as follows,

$$\langle Q\rangle_\psi(t) \equiv (\psi_S(t), Q_S\psi_S(t)) = (\psi_S(0), U(t)^\dagger Q_S U(t)\psi_S(0)) = (\psi_H, Q_H(t)\psi_H)\,, \tag{2.23}$$

where $\psi = \psi_S(0) = \psi_H$ is the state vector at $t = 0$ in the Schrödinger picture, and the time-independent state vector in the Heisenberg picture, while $Q = Q_S = Q_H(0)$ is the time-independent observable in the Schrödinger picture and the observable at $t = 0$ in the Heisenberg picture. The time-dependent quantities are $\psi_S(t) = U(t)\psi$ and $Q_H(t) = U(t)^\dagger Q U(t)$. Clearly, the expectation value of the observable is the same in both pictures at all times. From its definition, one sees that $Q_H(t)$ obeys the following equation of motion, $\dot{Q}(t) = dQ(t)/dt = i[H, Q(t)]$, so if $[H, Q] = 0$ then $[H, Q(t)] = 0$ and $Q(t) = Q$. Independently of what picture one is using, for conserved quantities one has $\langle Q\rangle_\psi(t) = \langle Q\rangle_\psi(0)$. Examples of continuous

[12]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 such that Eq. (2.20) holds, and that $\alpha = 0$ corresponds to the identity. See, e.g., Ref. [3].

transformations are translations and rotations, to which correspond respectively the conserved four-momentum and angular momentum.

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.

2.1 Free particles: 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. As a mtter of fact, the typical high-energy experiment consists in taking two bunch 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. This means that to a very good approximations these experiments involve free particles both in the initial and in the final state.

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. These observables constitute a complete set; energy is actually determined from the momenta through the dispersion relation $E^2 = \vec{p}^2 + m^2$. In denoting the state vector of a particle we usually put mass and the other defining observables under the label of the particle type, and further specify its momenta and third component of the spin. Using Dirac notation, we would write, e.g., for a neutral pion state $|\vec{p}; \pi^0\rangle$ (there is no spin here), for a proton state $|\vec{p}, s_z; p\rangle$, and so on. Other notations might be used, depending on what one wants to emphasise.

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

2.1.1 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 as $P\psi_{s_z}(\vec{x}) = \psi_{s_z}(-\vec{x})$, 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 in our reference frame, 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,

$$P|\vec{p}, s_z; \alpha\rangle = \eta_\alpha | - \vec{p}, s_z; \alpha\rangle \,, \tag{2.24}$$

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. If P is a symmetry, then a consistent assignment of phases in Eq. (2.24) can be

made. In the Heisenberg picture, in which the quantum states are given once and for all, and for all observers, the effect of symmetry transformations is shifted on the operators. In this picture one then finds for the momentum and spin *operators*

$$P^\dagger \vec{p} P = -\vec{p}, \qquad P^\dagger \vec{s} P = \vec{s}. \tag{2.25}$$

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 $PU(t) = U(t)P$: this implies $P^\dagger U(t) P = U(t)$, and so for infinitesimal t we find $P^\dagger i H P = iH$. 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.

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

$$\begin{aligned} \text{proton:} \qquad & \eta_p = \eta_p^{(0)} e^{i(\alpha+\gamma)} = 1\,, \\ \text{neutron:} \qquad & \eta_n = \eta_n^{(0)} e^{i\alpha} = 1\,, \\ \text{electron:} \qquad & \eta_e = \eta_e^{(0)} e^{i(\beta-\gamma)} = 1\,. \end{aligned} \tag{2.26}$$

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. 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.

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 $\eta_\alpha \eta_{\bar{\alpha}} = 1$, if they are bosons, and $\eta_\alpha \eta_{\bar{\alpha}} = -1$ if they are fermions. 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.

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

$$a\ b \to c\ d\,, \qquad\qquad a \to b\ c\,. \tag{2.27}$$

In a first approximation, we know that the transition probability for these processes are obtained from the interaction Hamiltonian (i.e., the full Hamiltonian H minus the free part H_0) via the relations

$$\text{transition probability } \propto\ |\langle c\ d|H_I|a\ b\rangle|^2\,, \quad |\langle b\ c|H_I|a\rangle|^2\,. \tag{2.28}$$

The very fact that a process happens implies that these matrix elements are nonzero.[13] Consider states with well-defined energy and orbital angular momentum, instead of momentum eigenstates. We know from quantum mechanics that for these states $P|\ell\ell_z\rangle = (-1)^\ell|\ell\ell_z\rangle$. If P is a symmetry of H, then since it is a symmetry of H_0 we have that $[P, H_I] = 0$, and so (omitting all irrelevant quantities from the notation)

$$\begin{aligned} 0 &= \langle \ell', \ell'_z; c\ d|[P, H_I]|\ell, \ell_z; a\ b\rangle = [(-1)^{\ell'}\eta_c\eta_d - (-1)^\ell\eta_a\eta_b]\langle \ell', \ell'_z; c\ d|H_I|\ell, \ell_z; a\ b\rangle\,, \\ 0 &= \langle \ell, \ell_z; b\ c|[P, H_I]|a\rangle = [(-1)^\ell\eta_b\eta_c - \eta_a]\langle \ell, \ell_z; b\ c|H_I|a\rangle\,, \end{aligned} \tag{2.29}$$

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 H_I are nonvanishing, Eq. (2.29) yields the relations

$$(-1)^{\ell'}\eta_c\eta_d = (-1)^\ell\eta_a\eta_b\,, \qquad (-1)^\ell\eta_b\eta_c = \eta_a\,. \tag{2.30}$$

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

$$\pi^-\ d \to n\ n\,, \tag{2.31}$$

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

[13] This is a simplification: processes could take place to higher orders in perturbation theory, so these matrix elements could vanish with the process still being possible. To get things straight we should replace H_I with the scattering operator S or the decay operator Γ, but the argument would be the same made here.

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.

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,

$$\begin{aligned} |\tfrac{1}{2}\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}\tfrac{1}{2}\rangle &= |11\rangle \\ \tfrac{1}{\sqrt{2}}\left(|\tfrac{1}{2}-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}\tfrac{1}{2}\rangle + |\tfrac{1}{2}\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}-\tfrac{1}{2}\rangle\right) &= |10\rangle \\ |\tfrac{1}{2}-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}-\tfrac{1}{2}\rangle &= |1-1\rangle \\ \tfrac{1}{\sqrt{2}}\left(|\tfrac{1}{2}-\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}\tfrac{1}{2}\rangle - |\tfrac{1}{2}\tfrac{1}{2}\rangle \otimes |\tfrac{1}{2}-\tfrac{1}{2}\rangle\right) &= |00\rangle\,, \end{aligned} \tag{2.32}$$

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 enconded 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 $\eta_\gamma = -1$.

2.1.2 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

$$C|\vec{p}, s_z; \alpha\rangle = \xi_\alpha |\vec{p}, s_z; \bar{\alpha}\rangle\,, \tag{2.33}$$

where again a phase ξ_α has to be included for generality. An argument identical to that used above for parity requires that C be a unitary operator and $[C, H] = 0$. Changing from particle to antiparticle leads to the change in sign of the internal quantum numbers, like, e.g., the electric charge q. 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. Other quantities that change sign under C are the baryon, lepton and lepton family numbers. For all these quantities, as well as for the electric charge, we have that $\{C, \mathcal{O}\} = 0$. If we reversed the discussion and started from this anticommutation relation 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 absolute value and opposite sign: indeed, these are the antiparticles. Applying C twice we find

$$C^2|\vec{p}, s_z; \alpha\rangle = \xi_\alpha \xi_{\bar{\alpha}} |\vec{p}, s_z; \alpha\rangle \,, \tag{2.34}$$

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).

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,

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} \propto \rho \,, \qquad & \vec{E} = -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \,, \\ \vec{\nabla} \wedge \vec{B} \propto \vec{J} \,, \qquad & \vec{B} = \vec{\nabla} \wedge \vec{A} \,. \end{aligned} \tag{2.35}$$

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 just a matter of convention: since they are not self-conjugate, there is no selection rule for them associated to 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 $\Gamma_{\pi^0 \to 3\gamma}/\Gamma_{\pi^0 \to 2\gamma} < 3.1 \cdot 10^{-8}$.

2.1.3 Time reversal and the CPT theorem

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

$$T|\vec{p}, s_z; \alpha\rangle = \zeta_{\alpha, s_z} |-\vec{p}, -s_z; \alpha\rangle \,, \tag{2.36}$$

where the phase ζ turns out to depend on s_z as well as on the particle species, and reads $\zeta_{\alpha,s_z} = (-1)^{s-s_z}\zeta_\alpha$. Contrary to P and C, T is an antiunitary symmetry. In fact, in contrast with Eq. (2.9), the requirement of invariance reads here $TU(t)\psi(0) = U(-t)T\psi(0)$, and so the operators must satisfy

$$TU(t)T^\dagger = U(-t) \Rightarrow TiHT^\dagger = -iH\,. \tag{2.37}$$

There are two ways to achieve this: if T is antilinear and antiunitary, then $Ti = -iT$ and we need $[T, H] = 0$, while if T is linear 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.

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:

$$\begin{aligned}\Theta|\vec{p}, s_z; \alpha\rangle &= CPT|\vec{p}, s_z; \alpha\rangle = CP\zeta_{\alpha,s_z}|-\vec{p}, -s_z; \alpha\rangle = C\eta_\alpha\zeta_{\alpha,s_z}|\vec{p}, -s_z; \alpha\rangle \\ &= \xi_\alpha\eta_\alpha\zeta_{\alpha,s_z}|\vec{p}, -s_z; \bar{\alpha}\rangle = \theta_{\alpha,s_z}|\vec{p}, -s_z; \bar{\alpha}\rangle\,,\end{aligned} \tag{2.38}$$

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

$$\begin{aligned}\Theta p^2|\vec{p}, s_z; \alpha\rangle &= m_\alpha^2\theta_{\alpha,s_z}|\vec{p}, -s_z; \bar{\alpha}\rangle \\ &= \Theta p^2\Theta^\dagger\Theta|\vec{p}, s_z; \alpha\rangle = p^2\theta_{\alpha,s_z}|\vec{p}, -s_z; \bar{\alpha}\rangle = m_{\bar{\alpha}}^2\theta_{\alpha,s_z}|\vec{p}, -s_z; \bar{\alpha}\rangle\,,\end{aligned} \tag{2.39}$$

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

$$\begin{aligned}\Gamma_\alpha &= \sum_f c_f|\langle f|H_I|\alpha\rangle|^2 = \sum_f c_f|\langle f|\Theta^\dagger H_I\Theta|\alpha\rangle|^2 \\ &= \sum_f c_f|\langle \bar{f}|H_I|\bar{\alpha}\rangle|^2 = \sum_f c_{\bar{f}}|\langle \bar{f}|H_I|\bar{\alpha}\rangle|^2 = \Gamma_{\bar{\alpha}}\,.\end{aligned} \tag{2.40}$$

2.2 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, we can assign to them the two-component vectors

$$p = \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \qquad n = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \tag{2.41}$$

The superposition principle then forces us to accept among the possible states of the nucleon any linear combination of these two,

$$N(\alpha, \beta) = \alpha p + \beta n = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} , \qquad \alpha, \beta \in \mathbb{C} . \tag{2.42}$$

What is assumed is that any of these states looks the same to the strong interactions. In mathematical terms, the assumed symmetry is under general unitary transformation of the nucleon state, i.e., under SU(2) transformations of the nucleon state. This is the *isospin* symmetry. The SU(2) group is a Lie group, with Lie algebra identical to that of the rotation group SO(3). In the case of isospin SU(2), the isospin generators are usually denoted as I_i, $i = 1, 2, 3$, and they satisfy the well-known commutation relations

$$[I_i, I_j] = i\varepsilon_{ijk} I_k . \tag{2.43}$$

Although not exact, isospin symmetry turns out to be a very good approximate symmetry of strong interactions, and we will soon see why.

Assuming that the strong Hamiltonian H_s is invariant under isospin rotations amounts to asking that $[\vec{I}, H_s] = 0$. This is also expressed by saying that isospin is conserved by strong interactions, since in the Heisenberg picture $\frac{d}{dt}\vec{I}(t) = i[H_s, \vec{I}(t)] = 0$. This assumption has two important consequences:

- $[\vec{I}, H_s] = 0$ implies that the spectrum of the theory is organised in degenerate isospin multiplets, which form the bases of irreducible representations of SU(2);
- the conservation law $\frac{d}{dt}\vec{I}(t) = 0$ implies that isospin is conserved in dynamical hadronic processes, i.e., in decay and scattering processes.

In hindsight, knowing of the existence of quarks and of QCD, we can trace isospin symmetry back to the symmetry under SU(2) rotations in the space of the up and down quarks, which (leaving

aside electromagnetic effects) is broken only by the small difference between their masses. In modern terms, we assign two-dimensional vectors to u and d as follows,

$$u = \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \qquad d = \begin{pmatrix} 0 \\ 1 \end{pmatrix} , \tag{2.44}$$

and find that the strong Hamiltonian is invariant under SU(2) rotations in this two-dimensional space, up to a symmetry-breaking term proportional to the mass difference $m_u - m_d$, which is much smaller than the typical strong interaction scale, i.e., 1 GeV. This is why isospin turns out to be such a good symmetry for strong interactions.

In the language of representation theory, the proton and the neutron, or the up and the down quark, form a basis of the *fundamental* (or *defining*) *representation* of SU(2). As is well known, irreducible representations of SU(2) are labelled by an integer or half-integer number I, which determines the total isospin $\vec{I}^2 = I(I+1)$ and the degeneracy $2I+1$ of the representation (see below). For the fundamental representation $I = \frac{1}{2}$. We will then use the notations

$$p = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |N; \tfrac{1}{2}\ \tfrac{1}{2}\rangle , \qquad n = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |N; \tfrac{1}{2}\ -\tfrac{1}{2}\rangle , \tag{2.45}$$

and

$$u = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |q; \tfrac{1}{2}\ \tfrac{1}{2}\rangle , \qquad d = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |q; \tfrac{1}{2}\ -\tfrac{1}{2}\rangle , \tag{2.46}$$

where the second argument is the value of I_3 associated to the particle.

In general, irreducible representations are most easily constructed making use of the raising and lowering operators $I_\pm = I_1 \pm iI_2$, which satisfy the commutation relations

$$[I_3, I_\pm] = \pm I_\pm , \qquad [\vec{I}^2, I_\pm] = 0 , \qquad [I_+, I_-] = 2I_3 . \tag{2.47}$$

Furthermore, one can show that

$$\vec{I}^2 = I_+I_- - I_3 + I_3^2 = I_-I_+ + I_3 + I_3^2 . \tag{2.48}$$

The first relation in Eq. (2.47) implies that if $|\psi\rangle$ is an eigenvector of $\vec{I}^2$ and I_3 with eigenvalues $C \geq 0$ and i_3, then $I_\pm|\psi\rangle$ is an eigenvector of $\vec{I}^2$ and I_3 with eigenvalues C and $i_3 \pm 1$. Let us focus on the raising operator. The norm of $I_+|\psi\rangle$ is $\langle\psi|I_-I_+|\psi\rangle = (C - i_3 - i_3^2)\langle\psi|\psi\rangle$, where we used Eq. (2.48). Since this must be a nonnegative quantity, after repeated application of I_+ we must find at some point that $I_+^{n+1}|\psi\rangle = 0$. Let I be the corresponding eigenvalue of I_3, and denote $|I\, I\rangle = I_+^n|\psi\rangle$, so that $I_+|I\, I\rangle = 0$. From the last relation we find that $C = I(I+1)$. Applying now I_- on this state we find vectors $I_-^k|I\, I\rangle = N_{I-k}|I\, I-k\rangle$ where N_{i_3} are normalisation factors. Taking all the $|I\, i_3\rangle$ to be normalised to 1 allows to determine recursively the N_{i_3} up to a phase via $|N_{i_3-1}|^2\langle I\, i_3 - 1|I\, i_3 - 1\rangle = |N_{i_3}|^2\langle I\, i_3|I_+I_-|I\, i_3\rangle$, i.e., $|N_{i_3-1}|^2 = |N_{i_3}|^2[I(I+1)+i_3-i_3^2]$, with $N_I = 1$. The phase of the vectors is arbitrary, and usually fixed according to the Condon-Shortley convention, asking for the raising and lowering operators $I_\pm = I_1 \pm iI_2$ have only positive matrix elements. In practice, $N_{i_3-1} = N_{i_3}\sqrt{I(I+1) + i_3 - i_3^2}$. Clearly $|N_{-I-1}|^2 = |N_{-I}|^2[I(I+1) - I - I^2] = 0$, so that $I_-|I\ -I\rangle = 0$, and the procedure stops after we have obtained a total of $2I+1$ vectors.

For future utility, we assign representative vectors to antiquarks as well,

$$\bar{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\bar{q}; \tfrac{1}{2}\ -\tfrac{1}{2}\rangle\,, \qquad \bar{d} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -|\bar{q}; \tfrac{1}{2}\ \tfrac{1}{2}\rangle\,. \tag{2.49}$$

These states form a basis of the *complex conjugate representation* of the group. There are three things to explain here: *i)* why is the sign of I_3 reversed, while *ii)* the two-dimensional representative vectors are the same for particle and antiparticle, and *iii)* why the minus sign in the isospin state corresponding to $\bar{d}$. The change in sign of I_3 is related to the fact that we are dealing with antiparticles, and as we will find out soon, I_3 is precisely one of those conserved charges that change sign under charge conjugation. At the same time, technical reasons in quantum field theory when implementing charge conjugation invariance, and to the phase convention $Cu = \bar{u}$ and $Cd = \bar{d}$, force us to use the very same vector representative for particle and antiparticle. Finally, this fact, combined with our will to use the Condon-Shortley convention for the isospin states $|I\, I_3\rangle$, forces us to introduce a minus sign in one of the two relations between isospin states and representative vectors.

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

$$H_{ho} = \hbar\omega(a_1^\dagger a_1 + a_2^\dagger a_2)\,. \tag{2.50}$$

The creation and annihilation operators satisfy the commutation relations

$$[a_i, a_j] = [a_i^\dagger, a_j^\dagger] = 0\,, \qquad [a_i, a_j^\dagger] = \delta_{ij}\,. \tag{2.51}$$

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

$$|n_1, n_2\rangle = \frac{1}{\sqrt{n_1!}}\frac{1}{\sqrt{n_2!}}(a_1^\dagger)^{n_1}(a_2^\dagger)^{n_2}|0\rangle\,, \tag{2.52}$$

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

$$a_2^\dagger a_1|n_1, n_2\rangle \propto |n_1 - 1, n_2 + 1\rangle\,, \qquad a_1^\dagger a_2|n_1, n_2\rangle \propto |n_1 + 1, n_2 - 1\rangle\,, \tag{2.53}$$

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., $a_i \to U_{ij}a_j$

particle	I	I_3	Q	B	S	Y
p	$\frac{1}{2}$	$\frac{1}{2}$	1	1	0	1
n	$\frac{1}{2}$	$-\frac{1}{2}$	0	1	0	1
π^+	1	1	1	0	0	0
π^0	1	0	0	0	0	0
π^-	1	-1	-1	0	0	0
K^+	$\frac{1}{2}$	$\frac{1}{2}$	1	0	1	1
K^0	$\frac{1}{2}$	$-\frac{1}{2}$	0	0	1	1
$\bar{K}^0$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	-1	-1
K^-	$\frac{1}{2}$	$-\frac{1}{2}$	-1	0	1	-1
Σ^+	1	1	1	1	-1	0
Σ^0	1	0	0	1	-1	0
Σ^-	1	-1	-1	1	-1	0
Δ^{++}	$\frac{3}{2}$	$\frac{3}{2}$	2	1	0	1
Δ^+	$\frac{3}{2}$	$\frac{1}{2}$	1	1	0	1
Δ^0	$\frac{3}{2}$	$-\frac{1}{2}$	0	1	0	1
Δ^-	$\frac{3}{2}$	$-\frac{3}{2}$	-1	1	0	1

Table 8: Isospin multiplets of hadrons.

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.

After this detour, let us get back to strong interactions, and discuss the implications of isospin symmetry for the hadron spectrum. As we mentioned above, this is expected to be organised in (approximately) degenerate multiplets corresponding to the irreducible representations of SU(2). As a matter of fact, besides the nucleon doublet formed by p and n, over the years several other multiplets were found, formed by hadrons with very close-by masses and identical baryon number and strangeness, but different electrical charge (pions, kaons, deltas, sigmas...). A total isospin I was assigned to these multiplets based on their degeneracy as degeneracy=$2I+1$, while within a multiplet states were distinguished by assigning an increasing value of I_3 to particles with increasing electric charge (empirically, no same-charge particles were found in multiplets). With this assignment of I_3, the following empirical formula for the electirc charge Q, known as Gell-Mann–Nishijima formula, applies:

$$Q = I_3 + \frac{1}{2}(B + S) = I_3 + \frac{1}{2}Y\,. \tag{2.54}$$

Here Q is the electric charge, B the baryon number and S the strangeness, and we have introduced for future utility the *hypercharge* $Y = B + S$. Examples of isospin multiplets are given in Table 8.

From a modern perspective, the classification of hadrons in isospin multiplets is easily understood in terms of quarks, and of the composition of fundamental and complex-conjugate representations of SU(2) (which are actually equivalent) and of the trivial representation. In

fact, from the assignments of Eqs. (2.46) and (2.49), and assigning $I = 0$ to the s quark and the corresponding antiquark, one can derive the mesonic multiplets of Table 8 by composing the two representations corresponding to the relevant quark and antiquark, and the baryonic multiplets by composing the three representations corresponding to the three relevant quarks. Here we will touch on the subject rather briefly, and we will discuss these points in greater detail when discussing the quark model. Since three quarks are needed to form a baryon, independently of the type of quarks involved, the baryon number of each quark is $\frac{1}{3}$. Explicitly, for a baryon made of n_u up quarks, n_d down quarks and n_s strange quarks one has

$$\left.\begin{array}{l} 1 = b_u n_u + b_d n_d + b_s n_s \\ 3 = n_u + n_d + n_s \end{array}\right\} \Longrightarrow \left(b_u - \frac{1}{3}\right) n_u + \left(b_d - \frac{1}{3}\right) n_d + \left(b_s - \frac{1}{3}\right) n_s = 0\,, \qquad (2.55)$$

for any n_u, n_d, n_s. Next, from the assignment of isospin to the quarks we have that $I_3 = \frac{1}{2}\,(n_u - n_d)$. We also associate strangeness to the presence of strange quarks, via $S = -n_s$. From this and from the isospin of the proton and neutron we find the quark content of these particles,

$$\left.\begin{array}{r} 1 = n_u^{(p)} - n_d^{(p)} = 2n_u^{(p)} - 3 \\ -1 = n_u^{(n)} - n_d^{(n)} = 2n_u^{(n)} - 3 \end{array}\right\} \Longrightarrow \quad n_u^{(p)} = 2\,, \quad n_u^{(n)} = 1\,. \qquad (2.56)$$

From the charge of p and n we find

$$\left.\begin{array}{l} 1 = q_u n_u^{(p)} + q_d n_d^{(p)} = 2q_u + q_d \\ 0 = q_u n_u^{(n)} + q_d n_d^{(n)} = q_u + 2q_d \end{array}\right\} \Longrightarrow \quad q_u = \frac{2}{3}\,, \quad q_d = -\frac{1}{3}\,. \qquad (2.57)$$

Using now the Lambda baryon, which is a neutral, isospin-singlet particle with strangeness $S = -1$, we find that $q_s = -\frac{1}{3}$. One can verify that in this way we fulfill the Gell-Mann–Nishijima at the level of quarks, so that it will be automatically satisfied at the level of hadrons. We can now give the expressions for the observables I_3, Q, B and S for any hadron in terms of the number of quarks and antiquarks: denoting with N_f the number of quarks of flavour f minus the number of antiquarks of the same flavour, $N_f = n_f - n_{\bar{f}}$, we have

$$I_3 = \tfrac{1}{2}(N_u - N_d)\,, \qquad Q = \tfrac{2}{3}N_u - \tfrac{1}{3}(N_d + N_s)\,, \qquad B = \tfrac{1}{3}(N_u + N_d + N_s)\,, \qquad S = -N_s\,. \quad (2.58)$$

The construction above seems a bit *ad hoc*, especially for what concerns the choice of the number of quarks and the assignment of isospin and strangeness. A full justification will come with the quark model.

It is now easy to build up the mesons containing only u and d by composing two $\frac{1}{2}$ representations: as is well known, $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$, so we get an isosinglet and an isotriplet. An isosinglet can also be obtained as $s\bar{s}$, and the two states may mix; we will not touch upon that state for the time being. The isotriplet, on the other hand, corresponds to the pions:

$$\begin{array}{rcccccl} -u\bar{d} & = & |q;\tfrac{1}{2}\,\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,\tfrac{1}{2}\rangle & = & |1\,1\rangle & = & \pi^+\,, \\ \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}) & = & \frac{1}{\sqrt{2}}\left(|q;\tfrac{1}{2}\,\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,-\tfrac{1}{2}\rangle + |q;\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,\tfrac{1}{2}\rangle\right) & = & |1\,0\rangle & = & \pi^0\,, \\ d\bar{u} & = & |q;\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \otimes |\bar{q};\tfrac{1}{2}\,-\tfrac{1}{2}\rangle & = & |1\,-1\rangle & = & \pi^-\,. \end{array} \qquad (2.59)$$

particle	I	I_3	Q	B	S	Y
u	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$
d	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$
s	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	-1	$-\frac{2}{3}$

Table 9: Isospin multiplets of quarks.

Kaons are obtained even more simply, since for them the relevant composition is $\frac{1}{2} \otimes 0 = \frac{1}{2}$. One then finds straightforwardly

$$\begin{aligned} u\bar{s} &= |q; \tfrac{1}{2}\ \tfrac{1}{2}\rangle \otimes |\bar{s}\rangle = |\tfrac{1}{2}\ \tfrac{1}{2}\rangle = K^+ , \\ d\bar{s} &= |q; \tfrac{1}{2}\ -\tfrac{1}{2}\rangle \otimes |\bar{s}\rangle = |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle = K^0 , \\ s\bar{u} &= |s\rangle \otimes |\bar{q}; \tfrac{1}{2}\ -\tfrac{1}{2}\rangle = |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle = K^- , \\ -s\bar{d} &= |s\rangle \otimes |\bar{q}; \tfrac{1}{2}\ \tfrac{1}{2}\rangle = |\tfrac{1}{2}\ \tfrac{1}{2}\rangle = \bar{K}^0 . \end{aligned} \tag{2.60}$$

Notice that the two multiplets are distinguished by strangeness. The goodness of isospin symmetry can be estimated by looking at mass splittings within the multiplets. We have for example

$$\begin{aligned} m_n &= 939.57 \text{ MeV} , & m_p &= 938.28 \text{ MeV} , \\ m_{\pi^\pm} &= 139.57 \text{ MeV} , & m_{\pi^0} &= 134.98 \text{ MeV} , \\ m_{K^\pm} &= 493.7 \text{ MeV} , & m_{K^0,\bar{K}^0} &= 497.6 \text{ MeV} , \end{aligned} \tag{2.61}$$

so that splittings are all of orders between the permille and the percent.

Besides static properties, isospin invariance has important consequences for dynamical processes as well. Consider for example the η meson, $m_\eta = 540$ 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,

$$R_2 |I\ I_3\rangle = (-1)^{I-I_3} |I\ -I_3\rangle . \tag{2.62}$$

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

$$\begin{aligned} G|\eta\rangle &= C|\eta\rangle = |\eta\rangle , \\ G|\pi^+\pi^-\pi^0\rangle &= C(-1)^{1-1}(-1)^{1+1}(-1)^{1-0}|\pi^-\pi^+\pi^0\rangle = -|\pi^+\pi^-\pi^0\rangle , \\ G|\pi^0\pi^0\pi^0\rangle &= C(-1)^3|\pi^0\pi^0\pi^0\rangle = -|\pi^0\pi^0\pi^0\rangle , \end{aligned} \tag{2.63}$$

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

$$\begin{aligned}
|\tfrac{3}{2}\ \tfrac{3}{2}\rangle &= |\tfrac{1}{2}\ \tfrac{1}{2}\rangle|1\ 1\rangle &&= |p\pi^+\rangle\,,\\
\sqrt{3}|\tfrac{3}{2}\ \tfrac{1}{2}\rangle &= |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle|1\ 1\rangle + \sqrt{2}|\tfrac{1}{2}\ \tfrac{1}{2}\rangle|1\ 0\rangle &&= |n\pi^+\rangle + \sqrt{2}|p\pi^0\rangle\,,\\
\sqrt{3}|\tfrac{3}{2}\ -\tfrac{1}{2}\rangle &= \sqrt{2}|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle|1\ 0\rangle + |\tfrac{1}{2}\ \tfrac{1}{2}\rangle|1\ -1\rangle &&= \sqrt{2}|n\pi^0\rangle + |p\pi^-\rangle\,,\\
|\tfrac{3}{2}\ -\tfrac{3}{2}\rangle &= |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle|1\ -1\rangle &&= |n\pi^-\rangle\,,\\
\sqrt{3}|\tfrac{1}{2}\ \tfrac{1}{2}\rangle &= \sqrt{2}|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle|1\ 1\rangle - |\tfrac{1}{2}\ \tfrac{1}{2}\rangle|1\ 0\rangle &&= \sqrt{2}|n\pi^+\rangle - |p\pi^0\rangle\,,\\
\sqrt{3}|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle &= |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle|1\ 0\rangle - \sqrt{2}|\tfrac{1}{2}\ \tfrac{1}{2}\rangle|1\ -1\rangle &&= |n\pi^0\rangle - \sqrt{2}|p\pi^-\rangle\,.
\end{aligned} \tag{2.64}$$

These relations can be inverted to achieve the desired decomposition. Here we will be concerned with the processes

$$p\ \pi^+ \to p\ \pi^+\,, \qquad p\ \pi^- \to p\ \pi^-\,, \qquad p\ \pi^- \to n\ \pi^0\,, \tag{2.65}$$

so we will need the following results,

$$\begin{aligned}
|p\pi^+\rangle &= |\tfrac{3}{2}\ \tfrac{3}{2}\rangle\,,\\
|p\pi^-\rangle &= \tfrac{1}{\sqrt{3}}\left(|\tfrac{3}{2}\ -\tfrac{1}{2}\rangle - \sqrt{2}|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle\right)\,,\\
|n\pi^0\rangle &= \tfrac{1}{\sqrt{3}}\left(\sqrt{2}|\tfrac{3}{2}\ -\tfrac{1}{2}\rangle + |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle\right)\,.
\end{aligned} \tag{2.66}$$

Exploiting invariance one can prove that

$$\begin{aligned}
\langle I' I'_3|S|II_3\rangle &= \delta_{I'I}\delta_{I'_3 I_3}\langle II_3|S|II_3\rangle\,,\\
\langle II_3|S|II_3\rangle &= \langle II'_3|S|II'_3\rangle \equiv \mathcal{M}_I\,.
\end{aligned} \tag{2.67}$$

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

$$\begin{aligned}
\mathcal{M}_{p\pi^+\to p\pi^+} &= \langle p\pi^+|S|p\pi^+\rangle = \langle\tfrac{3}{2}\ \tfrac{3}{2}|S|\tfrac{3}{2}\ \tfrac{3}{2}\rangle = \mathcal{M}_{\frac{3}{2}}\,,\\
\mathcal{M}_{p\pi^-\to p\pi^-} &= \langle p\pi^-|S|p\pi^-\rangle = \tfrac{1}{3}\left(\langle\tfrac{3}{2}\ -\tfrac{1}{2}|S|\tfrac{3}{2}\ -\tfrac{1}{2}\rangle + 2\langle\tfrac{1}{2}\ -\tfrac{1}{2}|S|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle\right) = \tfrac{1}{3}\left(\mathcal{M}_{\frac{3}{2}} + 2\mathcal{M}_{\frac{1}{2}}\right)\,,\\
\mathcal{M}_{p\pi^+\to n\pi^0} &= \langle n\pi^0|S|p\pi^-\rangle = \tfrac{\sqrt{2}}{3}\left(\langle\tfrac{3}{2}\ -\tfrac{1}{2}|S|\tfrac{3}{2}\ -\tfrac{1}{2}\rangle - \langle\tfrac{1}{2}\ -\tfrac{1}{2}|S|\tfrac{1}{2}\ -\tfrac{1}{2}\rangle\right) = \tfrac{\sqrt{2}}{3}\left(\mathcal{M}_{\frac{3}{2}} - \mathcal{M}_{\frac{1}{2}}\right)\,.
\end{aligned} \tag{2.68}$$

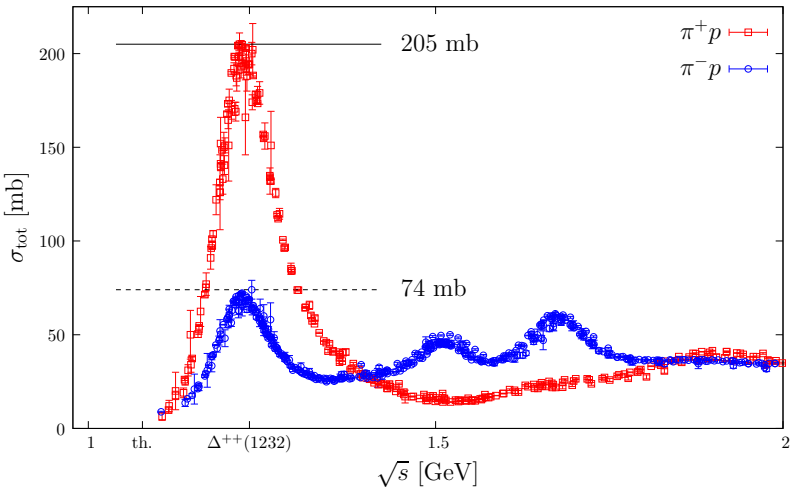

Figure 15: Total cross sections for $\pi^+ p$ and $\pi^- p$ scattering.

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 center-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 1.6). 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. (2.65) satisfy[14]

$$\begin{aligned} \frac{\sigma_{p\pi^+\to p\pi^+}}{\sigma_{p\pi^-\to p\pi^-}} &\simeq 9\,, \\ \frac{\sigma_{p\pi^+\to p\pi^+}}{\sigma_{p\pi^-\to n\pi^0}} &\simeq \frac{9}{2}\,. \end{aligned} \tag{2.69}$$

At this energy, the only scattering channel available for the $p\pi^+$ initial state is the elastic channel (there is enough energy for another neutral pion, but G-parity forbids its production), while for $p\pi^-$ there are the elastic channel and the inelastic process $p\pi^- \to n\pi^0$. We can then obtain the following relation for total cross sections,

$$\frac{\sigma_{p\pi^+\rm tot}}{\sigma_{p\pi^-\rm tot}} = \frac{\sigma_{p\pi^+\to p\pi^+}}{\sigma_{p\pi^-\to p\pi^-} + \sigma_{p\pi^-\to n\pi^0}} \underset{\sqrt{s}=m_\Delta}{\simeq} \frac{1}{\frac{1}{9}+\frac{2}{9}} = 3\,. \tag{2.70}$$

This relation is well verified experimentally (see Fig. 15).

A similar but easier calculation can be done to relate the three inelastic processes

$$pp \to d\pi^+\,, \quad pn \to d\pi^0\,, \quad nn \to d\pi^-\,. \tag{2.71}$$

[14]The proportionality factors between cross section and $|\mathcal{M}^2_{i\to f}$ 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.

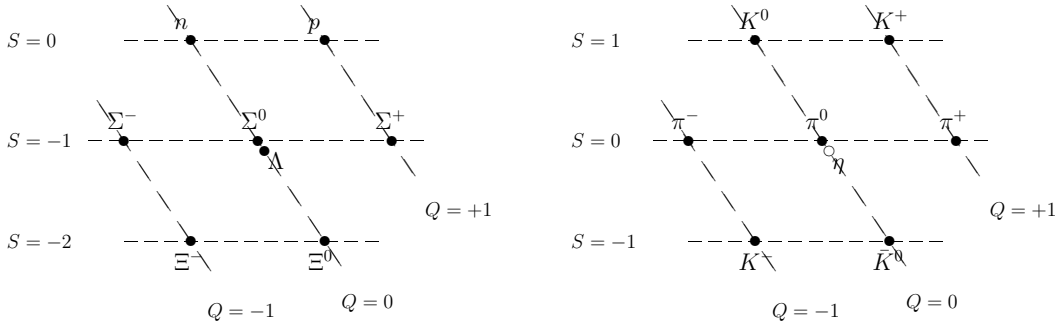

Figure 16: Lightest spin $\frac{1}{2}$ baryons (left) and lightest pseudoscalar mesons (right) known at the end of the '50s (black circles) organised in hexagonal patterns. Horizontal lines correspond to constant strangeness, diagonal lines correspond to constant electric charge. The η meson is shown as well with an empty circle.

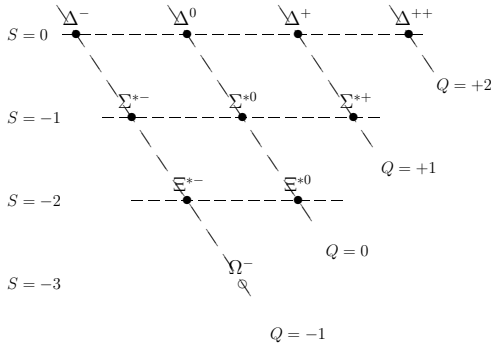

Figure 17: Spin-$\frac{3}{2}$ baryon resonances known at the end of the '50s (black circles) organised in a triangular pattern. Horizontal lines correspond to constant strangeness, diagonal lines correspond to constant electric charge. The Ω baryon is shown as well with an empty circle.

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

$$|pn\rangle = \frac{1}{\sqrt{2}}(|10\rangle + |00\rangle)\,. \tag{2.72}$$

One then concludes that $P_{pp\to d\pi^+} = P_{nn\to d\pi^-} = 2P_{pn\to d\pi^0}$.

Isospin symmetry is an example of an *internal symmetry*, i.e., a symmetry that acts on internal degrees of freedom, i.e., degrees of freedom unrelated to spacetime. In the next section we will discuss a bigger internal symmetry of strong interactions.

3 The quark model

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

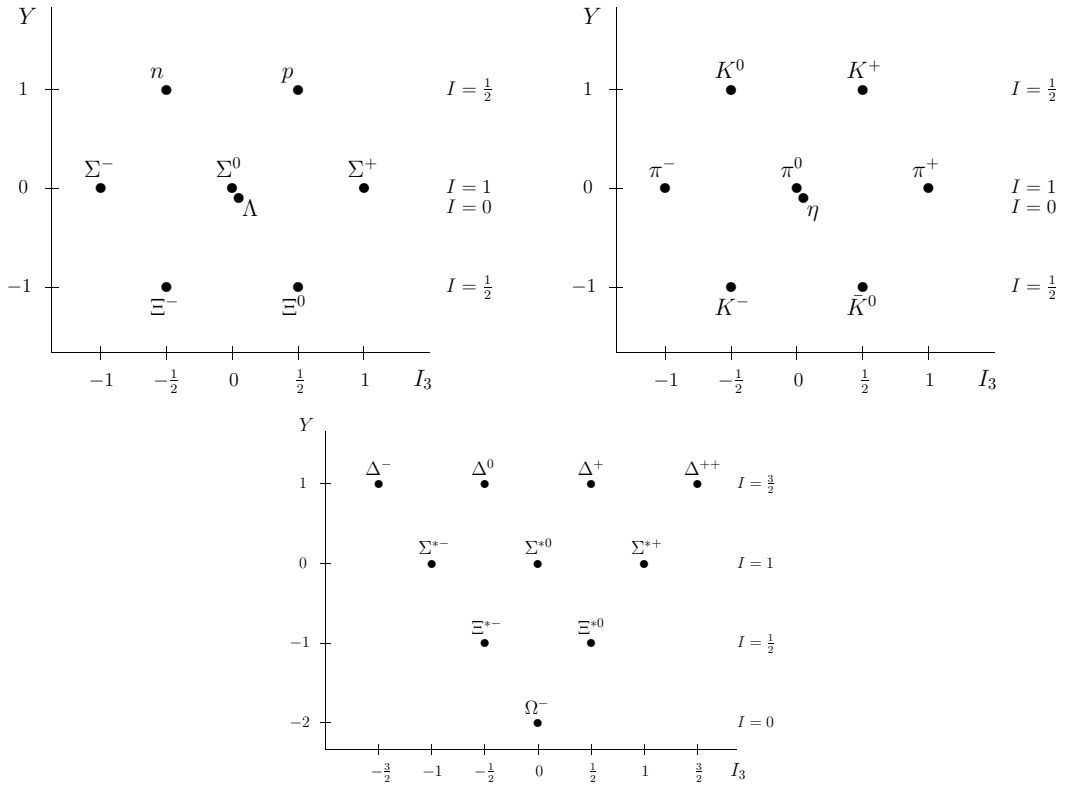

Figure 18: Baryons and mesons as in Figs. 16 and 17 but in the (I_3, Y) plane. Total isospin assignments are also shown.

above) and according to their strangeness.[15] This "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. 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. 16). 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. 17), 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

[15] "Strangeness" was literally associated to 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.

similar to that of the light baryons, with a "missing meson" in the center, and shifted by one unit in the direction of strangeness (see Fig. 16). Using instead the hypercharge $Y = B + S$, the two patterns would precisely overlap (see Fig. 18). Hypercharge seems therefore better suited than strangeness for a general classification of hadrons. Concerning mass splittings, they are still small within isospin mutiplets, but they are rather pronounced between mutiplets: being of the order of a few hundreds MeV, they are comparable with the meson masses. For the time being we will set aside this complication, and focus on the baryons.

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 $SU(2)_I$ isospin symmetry and the $U(1)_Y$ symmetry associated to 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_s were the sum of some exactly symmetric Hamiltonian, with degenerate multiplets of states corresponding to irreducible representations of some continuous internal symmetry group, and a symmetry-breaking term, $H_s = H_0 + H_I$. 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: for any $H_0|\psi\rangle = E|\psi\rangle$ we would have also $H_0\mathcal{O}_{a_1}\dots\mathcal{O}_{a_n}|\psi\rangle = \mathcal{O}_{a_1}\dots\mathcal{O}_{a_n}H_0|\psi\rangle = E\mathcal{O}_{a_1}\dots\mathcal{O}_{a_n}|\psi\rangle$. Not all these states are independent in general: finding out the dimension of the various irreducible representations is one of the main problems in representation theory. Since we are assuming the existence of an internal symmetry, the representations on the Hilbert space of the system should be unitary (we are not inverting time here).

3.1 Looking for a symmetry: the algebra of symmetry generators

If symmetry is the reason behind the hadron multiplets, then the symmetry generators responsible for the degeneracy of particle states in the limit of exact symmetry should allow to reproduce the patterns observed in Nature (e.g., the baryon octet) and to "navigate" within them. We already know a pretty accurate part of this symmetry, namely isospin symmetry, and we also know of strangeness and baryon number conservation: whatever the full symmetry group G is, it must contain $G \supset \mathrm{SU}(2)_I \times \mathrm{U}(1)_Y \times \mathrm{U}(1)_B$. This means that the isospin generators $\vec{I}$, the hypercharge Y and the baryon number B are among the symmetry generators. We know that the isospin generators commute with hypercharge, and both isospin and hypercharge commute with baryon number. On top of this, we know that I, I_3 and Y, besides the baryon number B, and spin $\vec{J}$, allow a full classification of the light hadrons. This means that $\{H, \vec{p}, J, J_3, B, I, I_3, Y\}$ is a complete set of commuting observables. Since we are looking for an internal symmetry, we will leave four-momentum and spin out of the discussion. Moreover, no degeneracy appears between baryons and mesons, so no symmetry generator should produce a change in baryon number. This means that all the other $\mathcal{O}_a$ must commmute with B, so in our search for G we can separate out the $\mathrm{U}(1)_B$ factor and ignore it – this is what we will do from now on. If we are to reproduce the observed multiplets, then given a simultaneous eigenvector of H_0 and of I_3 and Y with eigenvalues (i_3, y), it should be possible in general, using the symmetry generators, to construct eigenvectors with the same energy and with (I_3, Y)-eigenvalues equal to $(i_3 \pm \frac{1}{2}, y \pm 1)$ and $(i_3 \mp \frac{1}{2}, y \pm 1)$. This requires the introduction of new symmetry generators, beside isospin and hypercharge, and determines the commutation relations of these new operators with I_3 and Y.

We can now summarise our discussion by listing our assumptions:

1. we assume the existence of symmetry generators $\{\mathcal{O}_a\}$ such that the strong Hamiltonian splits into $H_s = H_0 + H_I$ with $[\mathcal{O}_a, H_0] = 0$;
2. we assume that $\{\mathcal{O}_a\}$ contains the Hermitian operators $\vec{I}$, Y, and four operators $V_\pm$ and $W_\pm$ that change (i_3, y) into $(i_3 \pm \frac{1}{2}, y \pm 1)$ and $(i_3 \mp \frac{1}{2}, y \pm 1)$;
3. we assume that no other symmetry generator exists (besides B);
4. we assume that there are no more than two symmetry generators that can be diagonalised simultaneously (again, besides B).

Assumption 3 has actually not been discussed above. We motivate it in the list below, where we also summarise the motivation for the other assumptions, as discussed above:

1. we want to explain the approximate degenerate multiplets in terms of an approximate continuous symmetry;
2. this symmetry must contain the already known symmetries, and should be able to reproduce the observed patterns in the (i_3, y) plane;
3. we want to keep the symmetry group minimal: the operators listed above are all it takes to move within the octet (and in the baryon resonance multiplet as well), so we hope that they will be enough;
4. isospin and hypercharge allow to fully classify hadrons (of given mass, spin and baryon number).

Before working out the consequences of our assumptions, it is worth recalling a few useful results. Let A and B be symmetry generators, $[A, H_0] = [B, H_0] = 0$. Then the following facts hold:

- any linear combination of A and B with complex coefficients is still a symmetry generator;
- the commutator $[A, B]$ is still a symmetry generator;
- the hermitian conjugates $A^\dagger$ and $B^\dagger$ are still symmetry generators.

The first result is quite trivial. The second result follows from the *Jacobi identity*: for any operators A_1, A_2, and A_3,

$$[[A_1, A_2], A_3] + [[A_3, A_1], A_2] + [[A_2, A_3], A_1] = 0\,, \tag{3.1}$$

and so

$$[[A, B], H_0] = -[[H_0, A], B] - [[B, H_0], A] = 0\,. \tag{3.2}$$

Finally, taking the Hermitian conjugate of $[H_0, A] = 0$ we find

$$0 = [H_0, A]^\dagger = -[H_0, A^\dagger]\,, \tag{3.3}$$

since $H_0 = H_0^\dagger$. The set of all symmetry generators is therefore closed under complex linear combinations, commutators and Hermitian conjugation. The first two properties make it a *complex Lie algebra*.

Let us now discuss in detail the consequences of our assumptions. Assumption 2 restates that $\vec{I}$ and Y are symmetry generators, and we already know everything there is to know about them. In particular, using the ladder operators $I_\pm = I_1 \pm iI_2$ we can recast their commutation relations $[I_i, I_j] = i\varepsilon_{ijk} I_k$ and $[\vec{I}, Y] = 0$ as

$$\begin{aligned} [I_3, Y] &= 0\,, \\ [I_3, I_\pm] &= \pm I_\pm\,, \qquad [Y, I_\pm] = 0\,, \\ [I_+, I_-] &= 2I_3\,. \end{aligned} \tag{3.4}$$

Assumption 2 tells us also about the existence of operators $V_\pm$ and $W_\pm$ that must satisfy

$$\begin{aligned} [I_3, V_\pm] &= \pm\frac{1}{2} V_\pm\,, \qquad [Y, V_\pm] = \pm V_\pm\,, \\ [I_3, W_\pm] &= \mp\frac{1}{2} W_\pm\,, \qquad [Y, W_\pm] = \pm W_\pm\,. \end{aligned} \tag{3.5}$$

These commutation relations follow from the requested action of these operators on eigenvectors: for example, from

$$I_3 V_+ |i_3, y\rangle = \left(i_3 + \tfrac{1}{2}\right) V_+ |i_3, y\rangle = (V_+ I_3 + [I_3, V_+])|i_3, y\rangle = i_3 V_+ |i_3, y\rangle + [I_3, V_+]|i_3, y\rangle\,, \tag{3.6}$$

that should hold on a complete set of states, we conclude $[I_3, V_+] = \frac{1}{2} V_+$. Finally, since I_3 and Y are assumed to be Hermitian, we have for example that

$$[I_3, V_+]^\dagger = \frac{1}{2} V_+^\dagger \Longrightarrow [I_3, V_+^\dagger] = -\frac{1}{2} V_+^\dagger\,, \tag{3.7}$$

i.e., given V_+, its conjugate $V_+^\dagger$ acts precisely as V_- should. Since, according to assumption 3, our set of symmetry generators is minimal, and since it must be closed under Hermitian conjugation, we have that V_- and $V_+^\dagger$ must be proportional. We can then choose

$$V_- \equiv V_+^\dagger\,, \qquad W_- \equiv W_+^\dagger\,. \tag{3.8}$$

As we will see now, assumptions 3 and 4 allow to fully determine the algebra of commutators (up to a sign that is fixed by the Hermiticity of the isospin generators). This is achieved more easily if we adopt a unified notation for the various symmetry generators. We will denote with $\vec{H} = (H_1, H_2)$ (not to be confused with the Hamiltonian) the operators

$$H_1 = I_3\,, \qquad H_2 = \kappa Y\,, \tag{3.9}$$

with κ a real positive scale factor to be determined along the way, and with $E_s^{(j)}$ the other operators,

$$E_\pm^{(1)} = I_\pm\,, \qquad E_\pm^{(2)} = V_\pm\,, \qquad E_\pm^{(3)} = W_\pm\,. \tag{3.10}$$

With this notation, the commutators Eqs. (3.4) and (3.5) read

$$\begin{aligned} [H_a, H_b] &= 0\,, \\ [\vec{H}, E_s^{(j)}] &= s\vec{\alpha}^{\,(j)} E_s^{(j)}\,, \end{aligned} \tag{3.11}$$

where

$$\vec{\alpha}^{\,(1)} = (1, 0)\,, \qquad \vec{\alpha}^{\,(2)} = (\tfrac{1}{2}, \kappa)\,, \qquad \vec{\alpha}^{\,(3)} = (-\tfrac{1}{2}, \kappa)\,. \tag{3.12}$$

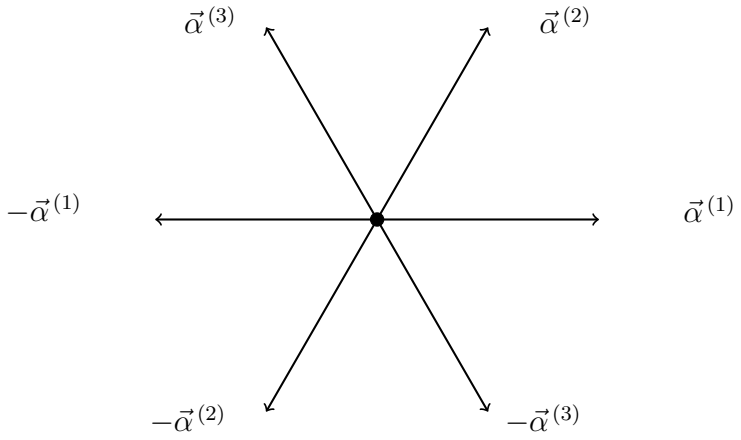

Figure 19: Root system of SU(3).

A convenient choice is $\kappa = \frac{\sqrt{3}}{2}$, so that $\vec{\alpha}^{(j)\,2} = 1\ \forall j$. In the language of Lie algebras, the vectors $\alpha^{(j)}$ are called the *root vectors* of the algebra, and form its *root system* (see Fig. 19). The commuting elements H_a form the *Cartan subalgebra*.

Eqs. (3.11) and (3.12) tell us an important fact. We can build a linear space V as the span of the symmetry generators, $V = \{\sum_j a_j H_j + \sum_{j,s} b_s^j E_s^{(j)} | a_j, b_s^j \in \mathbb{C}\}$. This linear space, together with the commutator, forms the Lie algebra associated with our symmetry generators. In this space, the action of a commutator with a fixed first argument is linear in the second, i.e., we can define on V the linear operator ad_X via $\mathrm{ad}_X Y \equiv [X, Y]$. Eqs. (3.11) and (3.12) then tell us that the $E_s^{(j)}$ are simultaneous eigenvectors of H_1 and H_2 with distinct pairs of nondegenerate eigenvalues. This makes these operators automatically linearly independent,[16] i.e., one can have $\sum_j a_j H_j + \sum_{j,s} b_s^j E_s^{(j)} = 0$ if and only if all the coefficients vanish, $a_j = b_s^j = 0$.

3.1.1 A direct determination of the commutators

We can now work out all the remaining commutators. From assumption 3 it follows that the commutator of any two operators must be a linear combination of them,

$$[E_s^{(j)}, E_t^{(k)}] = \vec{A}_{st}^{jk} \cdot \vec{H} + \sum_{l',u'} C_{stu'}^{jkl'} E_{-u'}^{(l')}\,, \tag{3.13}$$

where the minus sign in the subscript is chosen for convenience. Taking the commutator of the left-hand side with $\vec{H}$ and using the Jacobi identity we find

$$[\vec{H}, [E_s^{(j)}, E_t^{(k)}]] = -[E_t^{(k)}, [\vec{H}, E_s^{(j)}]] - [E_s^{(j)}, [E_t^{(k)}, \vec{H}]] = (s\vec{\alpha}^{(j)} + t\vec{\alpha}^{(k)})[E_s^{(j)}, E_t^{(k)}]\,, \tag{3.14}$$

i.e., $[E_s^{(j)}, E_t^{(k)}]$ is either an eigenvector of $\mathrm{ad}_{\vec{H}}$ or is zero. The first option is available only if $s\vec{\alpha}^{(j)} + t\vec{\alpha}^{(k)} = 0$, i.e., $k = j$ and $t = -s$, or, if $j \neq k$, if there are u, l such that

$$s\vec{\alpha}^{(j)} + t\vec{\alpha}^{(k)} + u\vec{\alpha}^{(l)} = 0\,. \tag{3.15}$$

[16] We know from the outset that H_1 and H_2 are independent from each other since they are not proportional, but suppose we did not know. Linear independence of H_1 and H_2 can be rephrased as the fact that there is no nonzero vector $\vec{\gamma}$ such that $\vec{\gamma} \cdot \vec{H} = 0$. If such a $\vec{\gamma}$ existed, then it should be simultaneously orthogonal to all the three $\vec{\alpha}^{(j)}$, which is clearly impossible.

In the first case $[E_s^{(j)}, E_{-s}^{(j)}]$ commutes with $\vec{H}$, and assumption 4 implies that it must be a linear combination thereof:

$$[E_s^{(j)}, E_{-s}^{(j)}] = 2s\vec{\beta}^{(j)} \cdot \vec{H}\,, \tag{3.16}$$

where the factor of 2 is again chosen for convenience. Notice that $\vec{\beta}^{(j)}$ must be real. Assumption 4 furthermore implies that $\vec{\beta}^{(j)} \neq \vec{0}$: if it were zero, then $E_s^{(j)}$, $E_{-s}^{(j)}$ and the linear combination $\vec{\alpha}_\perp^{(j)} \cdot \vec{H}$ with $\vec{\alpha}_\perp^{(j)} \cdot \vec{\alpha}^{(j)} = 0$ would form a set of three commuting operators, against our hypothesis that no such set of size bigger than 2 can be found. In the second case we have instead

$$[E_s^{(j)}, E_t^{(k)}] = C_{stu}^{jkl} E_{-u}^{(l)}\,, \tag{3.17}$$

where no sum is understood, and j, k, l and s, t, u are such that Eq. (3.15) is satisfied. Since

$$\vec{\alpha}^{(1)} - \vec{\alpha}^{(2)} + \vec{\alpha}^{(3)} = 0\,, \tag{3.18}$$

there are only six possible cases, corresponding to cyclic permutations of $(1+, 2-, 3+)$ and $(1-, 2+, 3-)$:

$$\begin{aligned}
&[E_+^{(1)}, E_-^{(2)}] = C_{+-+}^{1\,2\,3} E_-^{(3)}\,, \qquad [E_+^{(3)}, E_+^{(1)}] = C_{++-}^{3\,1\,2} E_+^{(2)}\,, \qquad [E_-^{(2)}, E_+^{(3)}] = C_{-++}^{2\,3\,1} E_-^{(1)}\,,\\
&[E_-^{(1)}, E_+^{(2)}] = C_{-+-}^{1\,2\,3} E_+^{(3)}\,, \qquad [E_-^{(3)}, E_-^{(1)}] = C_{--+}^{3\,1\,2} E_-^{(2)}\,, \qquad [E_+^{(2)}, E_-^{(3)}] = C_{+--}^{2\,3\,1} E_+^{(1)}\,.
\end{aligned} \tag{3.19}$$

Obviously, $C_{tsu}^{kjl} = -C_{stu}^{jkl}$. Moreover, since $E_s^{(j)} = E_{-s}^{(j)\dagger}$, we have that

$$C_{-s-t-u}^{jkl} = -C_{stu}^{jkl\,*}\,. \tag{3.20}$$

Before proceeding further, it is worth noting that a rescaling of the $E_s^{(j)}$ will not affect the vectors $\vec{\alpha}^{(j)}$, but it will affect both $\vec{\beta}^{(j)}$ and C_{stu}^{jkl}. In fact, these depend on the normalisation of our operators. One can easily show that under a redefinition $E_s^{(j)} \to \Lambda_s^{(j)} E_s^{(j)}$, with $\Lambda_{-s}^{(j)} = \Lambda_s^{(j)*}$, one finds

$$\begin{aligned}
\vec{\beta}^{(j)} &\to \frac{1}{|\Lambda_s^{(j)}|^2}\vec{\beta}^{(j)}\,,\\
C_{stu}^{jkl} &\to \frac{\Lambda_{-u}^{(l)}}{\Lambda_s^{(j)}\Lambda_t^{(k)}} C_{stu}^{jkl} = \frac{|\Lambda_u^{(l)}|^2}{\Lambda_s^{(j)}\Lambda_t^{(k)}\Lambda_u^{(l)}} C_{stu}^{jkl}\,.
\end{aligned} \tag{3.21}$$

We now take the commutator of Eq. (3.17) with $E_u^{(l)}$, again assuming that Eq. (3.15) is satisfied, and find

$$[E_u^{(l)}, [E_s^{(j)}, E_t^{(k)}]] = C_{stu}^{jkl}[E_u^{(l)}, E_{-u}^{(l)}] = C_{stu}^{jkl} 2u\vec{\beta}^{(l)} \cdot \vec{H}\,. \tag{3.22}$$

Summing over cyclic permutations of the indices we find

$$0 = C_{tus}^{klj} s\vec{\beta}^{(j)} + C_{ust}^{ljk} t\vec{\beta}^{(k)} + C_{stu}^{jkl} u\vec{\beta}^{(l)}\,, \tag{3.23}$$

where we have used the Jacobi identity and the linear independence of the H_a. Taking instead the commutator of Eq. (3.17) with $E_{-s}^{(j)}$ we find

$$\begin{aligned}
[E_{-s}^{(j)}, [E_s^{(j)}, E_t^{(k)}]] &= C_{stu}^{jkl}[E_{-s}^{(j)}, E_{-u}^{(l)}] = C_{stu}^{jkl} C_{-s-u-t}^{jlk} E_t^{(k)}\\
&= [E_s^{(j)}, [E_{-s}^{(j)}, E_t^{(k)}]] - [[E_s^{(j)}, E_{-s}^{(j)}], E_t^{(k)}] = -2st\vec{\beta}^{(j)} \cdot \vec{\alpha}^{(k)} E_t^{(k)}\,,
\end{aligned} \tag{3.24}$$

since $[E^{(j)}_{-s}, E^{(k)}_{t}] = 0$. We then have the identities

$$C^{jkl}_{stu} C^{jlk}_{-s-u-t} = -2st \vec{\beta}^{\,(j)} \cdot \vec{\alpha}^{\,(k)} \,. \tag{3.25}$$

Permuting k, t with l, u, and changing the sign of s, t, u, we get

$$C^{jlk}_{-s-u-t} C^{jkl}_{stu} = -2su \vec{\beta}^{\,(j)} \cdot \vec{\alpha}^{\,(l)} \,, \tag{3.26}$$

with the same left-hand side. Summing and subtracting these two identities we obtain

$$\begin{aligned} C^{jkl}_{stu} C^{jlk}_{-s-u-t} = -s\vec{\beta}^{\,(j)} \cdot (t\vec{\alpha}^{\,(k)} + u\vec{\alpha}^{\,(l)}) = \vec{\beta}^{\,(j)} \cdot \vec{\alpha}^{\,(j)} \,, \\ 0 = \vec{\beta}^{\,(j)} \cdot (t\vec{\alpha}^{\,(k)} - u\vec{\alpha}^{\,(l)}) \,. \end{aligned} \tag{3.27}$$

The second equation in Eq. (3.27) determines the $\vec{\beta}^{\,(j)}$ up to a real factor:

$$\begin{aligned} &\vec{\beta}^{\,(1)} \perp \vec{\alpha}^{\,(2)} + \vec{\alpha}^{\,(3)} = (0, \sqrt{3}) \qquad &\Rightarrow \qquad &\vec{\beta}^{\,(1)} = N^{(1)}(1,0) = N^{(1)} \vec{\alpha}^{\,(1)} \,, \\ &\vec{\beta}^{\,(2)} \perp \vec{\alpha}^{\,(1)} - \vec{\alpha}^{\,(3)} = (\tfrac{3}{2}, -\tfrac{\sqrt{3}}{2}) \qquad &\Rightarrow \qquad &\vec{\beta}^{\,(2)} = N^{(2)}(\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}) = N^{(2)} \vec{\alpha}^{\,(2)} \,, \\ &\vec{\beta}^{\,(3)} \perp \vec{\alpha}^{\,(1)} + \vec{\alpha}^{\,(2)} = (\tfrac{3}{2}, \tfrac{\sqrt{3}}{2}) \qquad &\Rightarrow \qquad &\vec{\beta}^{\,(3)} = N^{(3)}(-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}) = N^{(3)} \vec{\alpha}^{\,(3)} \,. \end{aligned} \tag{3.28}$$

We already know that $\vec{\beta}^{\,(1)} = \vec{\alpha}^{\,(1)}$, so that $N^{(1)} = 1$: what the first of the equations above tells us is that this is compatible with our construction. Plugging Eq. (3.28) into Eq. (3.23) we find

$$0 = C^{klj}_{tus} N^{(j)} s \vec{\alpha}^{\,(j)} + C^{ljk}_{ust} N^{(k)} t \vec{\alpha}^{\,(k)} + C^{jkl}_{stu} N^{(l)} u \vec{\alpha}^{\,(l)} \,, \tag{3.29}$$

which together with Eq. (3.18) implies that[17]

$$C^{klj}_{tus} N^{(j)} = C^{ljk}_{ust} N^{(k)} = C^{jkl}_{stu} N^{(l)} \,. \tag{3.30}$$

Plugging Eq. (3.28) into the first equation in Eq. (3.27) yields

$$C^{jkl}_{stu} C^{jlk}_{-s-u-t} = N^{(j)} \,. \tag{3.31}$$

Multiplying both sides by $N^{(k)} N^{(l)}$, using the cyclicity implied by Eq. (3.30) and the relation Eq. (3.20), as well as antisymmetry in the first two indices, we find

$$\begin{aligned} N^{(j)} N^{(k)} N^{(l)} &= C^{jkl}_{stu} N^{(l)} C^{jlk}_{-s-u-t} N^{(k)} = -C^{jkl}_{stu} N^{(l)} C^{ljk}_{-u-s-t} N^{(k)} \\ &= -C^{jkl}_{stu} N^{(l)} C^{jkl}_{-s-t-u} N^{(l)} = |C^{jkl}_{stu}|^2 N^{(l)2} \,. \end{aligned} \tag{3.32}$$

Setting $N^{(j)} = \bar{N}^{(j)} (-1)^{\sigma^{(j)}}$, with $\sigma^{(j)} = 0$ or 1 and $\bar{N}^{(j)} > 0$, this gives

$$(-1)^{\sigma^{(j)} + \sigma^{(k)} + \sigma^{(k)}} = 1 \,, \qquad |C^{jkl}_{stu}|^2 = \frac{\bar{N}^{(j)} \bar{N}^{(k)}}{\bar{N}^{(l)}} \,. \tag{3.33}$$

[17] If the vectors $\vec{\alpha}^{\,(j)}$ satisfy two different equations, $\sum_{j=1}^{3} a_j \vec{\alpha}^{\,(j)} = 0$ and $\sum_{j=1}^{3} b_j \vec{\alpha}^{\,(j)} = 0$, they must do so with either all vanishing coefficient or all nonvanishing coefficients, since they are pairwise linearly independent vectors. Multiplying the first equation by b_3 and the second one by a_3 and subtracting we find $\sum_{j=1}^{2} (b_3 a_j - a_3 b_j) \vec{\alpha}^{\,(j)} = 0$, which can be true only if $b_3 a_j = a_3 b_j \ \forall j$, i.e., if $b_j \propto a_j$.

Since $N^{(1)} = 1$, i.e., $\bar{N}^{(1)} = 1$ and $\sigma^{(1)} = 0$, this relation implies that $N^{(2)}$ and $N^{(3)}$ must have the same sign, i.e., $\sigma^{(2)} = \sigma^{(3)}$, yet to be determined. However, we will act as if we did not know that: from this we could still conclude that there can be either none or two negative $N^{(j)}$. From Eq. (3.30) we find instead

$$C^{klj}_{tus}\bar{N}^{(j)}(-1)^{\sigma^{(j)}} = C^{ljk}_{ust}\bar{N}^{(k)}(-1)^{\sigma^{(k)}} = C^{jkl}_{stu}\bar{N}^{(l)}(-1)^{\sigma^{(l)}}\,. \tag{3.34}$$

It is easy to see, using Eq. (3.20), that the key relations Eqs. (3.33) and (3.34) remain unchanged under redefinitions of our operators. We can then make a redefinition with $\Lambda^{(j)}_s = \sqrt{\bar{N}^{(j)}}$ to get them in the form

$$\begin{aligned} &(-1)^{\sigma^{(j)}+\sigma^{(k)}+\sigma^{(k)}} = 1\,, \qquad |C^{jkl}_{stu}|^2 = 1\,,\\ &C^{klj}_{tus}(-1)^{\sigma^{(j)}} = C^{ljk}_{ust}(-1)^{\sigma^{(k)}} = C^{jkl}_{stu}(-1)^{\sigma^{(l)}}\,, \end{aligned} \tag{3.35}$$

in terms of the new coefficients. The C^{jkl}_{stu} are now just phases. Under a further redefinition with $\Lambda^{(j)}_s = e^{is\delta^{(j)}}$ these relations do not change, but $C^{jkl}_{stu} \to C^{jkl}_{stu}e^{-i(s\delta^{(j)}+t\delta^{(k)}+u\delta^{(l)})}$, so we can fix one of them to 1, and the corresponding coefficient with $stu \to -s-t-u$ to -1. All that remains to be done is fixing the $\sigma^{(j)}$.

For this last step, recall that with our normalisation

$$[E^{(j)}_s, E^{(k)}_t] = e^{i\Phi^{jkl}_{stu}}E^{(l)}_{-u}\,. \tag{3.36}$$

Multiplying on the left by $E^{(l)}_u$ we find

$$\begin{aligned} e^{i\Phi^{jkl}_{stu}}E^{(l)}_u E^{(l)}_{-u} &= E^{(l)}_u(E^{(j)}_s E^{(k)}_t - E^{(k)}_t E^{(j)}_s) = [E^{(l)}_u, E^{(j)}_s]E^{(k)}_t + [E^{(j)}_s, E^{(l)}_u E^{(k)}_t]\\ &= e^{i\Phi^{ljk}_{ust}}E^{(k)}_{-t}E^{(k)}_t + [E^{(j)}_s, E^{(l)}_u E^{(k)}_t]\,. \end{aligned} \tag{3.37}$$

Restricting to a fixed energy subspace and taking the trace on both sides, the commutator term gives no contribution, and

$$e^{i\Phi^{jkl}_{stu}}\tilde{\mathrm{tr}}\, E^{(l)}_u E^{(l)}_{-u} = e^{i\Phi^{ljk}_{ust}}\tilde{\mathrm{tr}}\, E^{(k)}_{-t}E^{(k)}_t\,, \tag{3.38}$$

where $\tilde{\mathrm{tr}}$ denotes our restricted way of taking the trace. But $E^{(l)}_u = E^{(l)\,\dagger}_{-u}$, so both traces are positive real numbers, and we conclude $e^{i\Phi^{jkl}_{stu}} = e^{i\Phi^{ljk}_{ust}} = e^{i\Phi^{klj}_{tus}}$. This implies that

$$(-1)^{\sigma^{(j)}} = (-1)^{\sigma^{(k)}} = (-1)^{\sigma^{(l)}}\,, \tag{3.39}$$

which together with the first equation in Eq. (3.35) implies $(-1)^{3\sigma^{(j)}} = 1$, i.e.,

$$\sigma^{(j)} = \sigma^{(k)} = \sigma^{(l)} = 0\,. \tag{3.40}$$

From this it also follows that

$$C^{klj}_{tus} = C^{ljk}_{ust} = C^{jkl}_{stu}\,. \tag{3.41}$$

We can finally use our freedom to redefine the phase of the operators to set

$$C^{2\,3\,1}_{+--} = C^{3\,1\,2}_{--+} = C^{1\,2\,3}_{-+-} = 1\,, \qquad C^{2\,3\,1}_{-++} = C^{3\,1\,2}_{++-} = C^{1\,2\,3}_{+-+} = -1\,. \tag{3.42}$$

Summarising, besides Eqs. (3.4) and (3.5), the only nonzero commutators are the following:

$$\begin{aligned} &[V_+, W_-] = I_+\,, \qquad [I_-, V_+] = W_+\,, \qquad [W_-, I_-] = V_-\,, \\ &[V_-, W_+] = -I_-\,, \qquad [I_+, V_-] = -W_-\,, \qquad [W_+, I_+] = -V_+\,, \end{aligned} \tag{3.43}$$

together with

$$[E_+^{(j)}, E_-^{(j)}] = 2\vec{\alpha}^{\,(j)} \cdot \vec{H} \qquad \Longrightarrow \qquad \begin{cases} [I_+, I_-] = 2H_1 = 2I_3\,, \\ [V_+, V_-] = H_1 + \sqrt{3}H_2 = I_3 + \frac{3}{2}Y\,, \\ [W_+, W_-] = -H_1 + \sqrt{3}H_2 = -I_3 + \frac{3}{2}Y\,. \end{cases} \tag{3.44}$$

We can replace our set of symmetry generators with one made entirely of Hermitian operators. This simply requires to trade $I_\pm$, $V_\pm$ and $W_\pm$ with

$$\begin{aligned} I_1 &= \frac{I_+ + I_-}{2}\,, & I_2 &= \frac{I_+ - I_-}{2i}\,, \\ V_1 &= \frac{V_+ + V_-}{2}\,, & V_2 &= \frac{V_+ - V_-}{2i}\,, \\ W_1 &= \frac{W_+ + W_-}{2}\,, & W_2 &= \frac{W_+ - W_-}{2i}\,. \end{aligned} \tag{3.45}$$

The commutation relations of these operators are obtained from Eqs. (3.5), (3.43) and (3.44), together with the fact that all other commutators vanish. From Eq. (3.5) we get

$$\begin{aligned} &[\vec{H}, I_1] = i\vec{\alpha}^{\,(1)} I_2\,, \qquad [\vec{H}, I_2] = -i\vec{\alpha}^{\,(1)} I_1\,, \\ &[\vec{H}, V_1] = i\vec{\alpha}^{\,(2)} V_2\,, \qquad [\vec{H}, V_2] = -i\vec{\alpha}^{\,(2)} V_1\,, \\ &[\vec{H}, W_1] = i\vec{\alpha}^{\,(3)} W_2\,, \qquad [\vec{H}, W_2] = -i\vec{\alpha}^{\,(3)} W_1\,, \end{aligned} \tag{3.46}$$

while from Eq. (3.43) a little patience leads to

$$\begin{aligned} [V_1, W_1] &= [V_2, W_2] = \tfrac{i}{2} I_2\,, \qquad [V_1, W_2] = -[V_2, W_1] = \tfrac{i}{2} I_1\,, \\ [I_1, V_1] &= [I_2, V_2] = \tfrac{i}{2} W_2\,, \qquad [I_1, V_2] = -[V_2, I_1] = \tfrac{i}{2} W_1\,, \\ [W_1, I_1] &= -[W_2, I_2] = -\tfrac{i}{2} V_2\,, \qquad [W_1, I_2] = [W_2, I_1] = \tfrac{i}{2} V_1\,. \end{aligned} \tag{3.47}$$

Finally, from Eq. (3.44) we obtain

$$\begin{aligned} [I_1, I_2] &= i I_3\,, \\ [V_1, V_2] &= i(\tfrac{1}{2} I_3 + \tfrac{3}{4} Y)\,, \\ [W_1, W_2] &= i(-\tfrac{1}{2} I_3 + \tfrac{3}{4} Y)\,. \end{aligned} \tag{3.48}$$

Eqs. (3.46) to (3.48) show an important fact: it is possible to express the algebra of commutators in the form

$$[\mathcal{O}_a, \mathcal{O}_b] = i f_{abc} \mathcal{O}_c\,, \tag{3.49}$$

where $\{\mathcal{O}_a\} = \{I_{1,2,3}, V_{1,2}, W_{1,2}, \frac{\sqrt{3}}{2} Y\}$, with a running on $\{1, \ldots, 8\}$ in the given order, and where f_{abc} are *real* coefficients. This indicates that the $\{\mathcal{O}_a\}$ generate a real Lie algebra with

structure constants given by f_{abc}. One can prove the following important properties of the structure constants in the case at hand:

$$f_{abc} = -f_{bac}\,, \quad f_{abc} = f_{cab} = f_{bca}\,, \quad f_{abc}f_{abd} = 3\delta_{cd}\,. \tag{3.50}$$

The first property is clearly general, as it originates in the antisymmetric nature of the commutator. The second one is true for any compact semi-simple Lie algebra, i.e., for the Lie algebra of any compact semi-simple Lie group.[18] Finally, the third property depends on our choice of normalisation of the various operators, although positive-semidefiniteness of the quadratic form $g_{cd} \equiv f_{abc}f_{abd}$ holds in general for compact semi-simple Lie algebras; conversely, positive-definiteness of g_{ab} implies that the Lie algebra is compact and semi-simple. This is reassuring: since we are looking for a symmetry group that admits finite-dimensional unitary representations (i.e., the hadron multiplets), it better be compact. In the present context, the identities of Eq. (3.49) can be derived by direct inspection: we will se below a more elegant way.

3.1.2 A smarter way to determine the commutators: the Killing form

The approach used above to determine the commutators was very much hands-down. A more elegant approach, which also leads to an easy proof of the identities in Eq. (3.50), is based on the so-called *Killing form*. Everything proceeds as above up to Eqs. (3.15)–(3.17). We then make use of the linear operators ad_X to define a bilinear form on our algebra:

$$B(X,Y) \equiv \mathrm{tr}\,\mathrm{ad}_X\mathrm{ad}_y\,. \tag{3.51}$$

It helps here recalling what is the intrinsic meaning of the trace. Given an operator A defined on a linear space V spanned by a basis $\{e_n\}$, we have that Ae_n can be decomposed as a linear combination of basis vectors, $Ae_n = \sum_m c_{nm}e_m$. The trace is then $\mathrm{tr}\,A = \sum_n c_{nn}$. One can show that $\mathrm{tr}\,A$ does not depend on the choice of basis. In the case at hand, to compute $B(X,Y)$ we apply $\mathrm{ad}_X\mathrm{ad}_y$ on the generators $\vec{H}$ and $E_s^{(j)}$ of our algebra, decompose the result, look for the appropriate coefficient and perform the sum.

The bilinear form $B(X,Y)$ is the Killing form of the algebra, and has an important property that follows from the following identity:

$$\begin{aligned}\mathrm{ad}_{[X,Y]}Z = [[X,Y],Z] &= -[[Z,X],Y] - [[Y,Z],X] = [X,[Y,Z]] - [Y,[X,Z]] \\ &= [\mathrm{ad}_X, \mathrm{ad}_Y]Z\,,\end{aligned} \tag{3.52}$$

i.e., $\mathrm{ad}_{[X,Y]} = [\mathrm{ad}_X, \mathrm{ad}_Y]$. It then follows that

$$\begin{aligned}B([X,Y],Z) &= \mathrm{tr}\,\mathrm{ad}_{[X,Y]}\mathrm{ad}_Z = \mathrm{tr}\,[\mathrm{ad}_X, \mathrm{ad}_Y]\mathrm{ad}_Z \\ &= \mathrm{tr}\,\mathrm{ad}_X[\mathrm{ad}_Y, \mathrm{ad}_Z] = -\mathrm{tr}\,[\mathrm{ad}_X, \mathrm{ad}_Z]\mathrm{ad}_Y \\ &= B(X,[Y,Z]) = -B([X,Z],Y)\,,\end{aligned} \tag{3.53}$$

or

$$B([X,Y],Z) = B([Y,Z],X) = B([Z,X],Y)\,. \tag{3.54}$$

[18] A simple algebra is a non-Abelian algebra that does not contain non-trivial ideals. An ideal $\mathfrak{a} \subset \mathfrak{g}$ is a linear subspace of the algebra $\mathfrak{g}$ that is invariant under commutators with any element of the algebra, $[\mathfrak{g}, \mathfrak{a}] \subset \mathfrak{a}$; it is then also a subalgebra of $\mathfrak{g}$. An algebra is semi-simple if it is the direct sum of simple algebras, i.e, the direct sum of simple algebras commuting with each other.

This property allows us to greatly reduce the number of entries that we have to compute explicitly. As a first consequence of Eq. (3.53), we have that $B(E_s^{(j)}, \vec{H}) = 0$. In fact,

$$\begin{aligned} B([s\vec{\alpha}^{(j)} \cdot \vec{H}, E_s^{(j)}], H_b) = -B(E_s^{(j)}, [s\vec{\alpha}^{(j)} \cdot \vec{H}, H_b]) = 0\,, \\ \vec{\alpha}^{(j)2} B(E_s^{(j)}, H_b) = B(E_s^{(j)}, H_b) = 0\,. \end{aligned} \tag{3.55}$$

Furthermore,

$$0 = B([\vec{H}, E_s^{(j)}], E_t^{(k)}) + B([\vec{H}, E_t^{(k)}], E_s^{(j)}) = (s\vec{\alpha}^{(j)} + t\vec{\alpha}^{(k)}) B(E_s^{(j)}, E_t^{(k)})\,, \tag{3.56}$$

which implies that $B(E_s^{(j)}, E_t^{(k)}) = 0$ unless $k = j$ and $t = -s$. In this case Eq. (3.53) implies

$$\begin{aligned} B([\vec{H}, E_s^{(j)}], E_{-s}^{(j)}) = B(\vec{H}, [E_s^{(j)}, E_{-s}^{(j)}])\,, \\ s\vec{\alpha}^{(j)} B(E_s^{(j)}, E_{-s}^{(j)}) = 2sB(\vec{H}, \vec{H} \cdot \vec{\beta}^{(j)})\,, \\ \alpha_a^{(j)} B(E_s^{(j)}, E_{-s}^{(j)}) = 2B(H_a, H_b)\beta_b^{(j)}\,, \end{aligned} \tag{3.57}$$

where sum over the repeated index b is understood. The elements $B(H_a, H_b)$ have to be calculated explicitly, but this is easy. Let us introduce the following notation: if A is an element of a linear space that decomposes as $A = \sum_n c_n e_n$ in the basis $\{e_n\}$, then we write $\mathrm{coeff}_{e_n} A = c_n$. Then

$$\begin{aligned} B(H_a, H_b) = \mathrm{tr}\, \mathrm{ad}_{H_a} \mathrm{ad}_{H_b} &= \sum_c \mathrm{coeff}_{H_c} [H_a, [H_b, H_c]] + \sum_{j,s} \mathrm{coeff}_{E_s^{(j)}} [H_a, [H_b, E_s^{(j)}]] \\ &= \sum_{j,s} \mathrm{coeff}_{E_s^{(j)}} s^2 \alpha_a^{(j)} \alpha_b^{(j)} E_s^{(j)} = 2 \sum_j \alpha_a^{(j)} \alpha_b^{(j)}\,. \end{aligned} \tag{3.58}$$

Explicitly,

$$\begin{aligned} B(H_1, H_1) = 2 \sum_j \alpha_1^{(j)} \alpha_1^{(j)} = 2\left[1 \cdot 1 + \tfrac{1}{2} \cdot \tfrac{1}{2} + \left(-\tfrac{1}{2}\right) \cdot \left(-\tfrac{1}{2}\right)\right] = 3\,, \\ B(H_2, H_2) = 2\left[0 \cdot 0 + \tfrac{\sqrt{3}}{2} \cdot \tfrac{\sqrt{3}}{2} + \tfrac{\sqrt{3}}{2} \cdot \tfrac{\sqrt{3}}{2}\right] = 3\,, \\ B(H_1, H_2) = 2\left[1 \cdot 0 + \tfrac{1}{2} \cdot \tfrac{\sqrt{3}}{2} + \left(-\tfrac{1}{2}\right) \cdot \tfrac{\sqrt{3}}{2}\right] = 0\,, \end{aligned} \tag{3.59}$$

and so

$$B(H_a, H_b) = 3\delta_{ab}\,, \quad a, b = 1, 2\,. \tag{3.60}$$

Notice that if we had not fixed κ yet in Eq. (3.12), then

$$\begin{aligned} B(H_2, H_2) = 2 \sum_j \alpha_2^{(j)} \alpha_2^{(j)} = 2\left[0 \cdot 0 + \kappa \cdot \kappa + \kappa \cdot \kappa\right] = 4\kappa^2\,, \\ B(H_1, H_2) = 2 \sum_j \alpha_2^{(j)} \alpha_2^{(j)} = 2\left[1 \cdot 0 + \tfrac{1}{2} \cdot \kappa + \left(-\tfrac{1}{2}\right) \cdot \kappa\right] = 0\,, \end{aligned} \tag{3.61}$$

and we would choose $\kappa = \frac{\sqrt{3}}{2}$ to obtain the diagonal form of Eq. (3.60). The only other nonzero elements of the Killing form are then determined by the relation

$$\alpha_a^{(j)} B(E_s^{(j)}, E_{-s}^{(j)}) = 2B(H_a, H_b)\beta_b^{(j)} = 2B(H_a, H_a)\beta_a^{(j)}\,, \tag{3.62}$$

which, since $\vec{\beta}^{(j)} \neq \vec{0}$, implies that $B(E_s^{(j)}, E_{-s}^{(j)})$ is nonzero. We are free to choose $|B(E_s^{(j)}, E_{-s}^{(j)})|$ as we please, since we can always rescale the operators (which are however related, $E_{-s}^{(j)} = E_s^{(j)\dagger}$). We choose

$$|B(E_s^{(j)}, E_{-s}^{(j)})| = 2B(H_a, H_a) = 6\,, \tag{3.63}$$

independently of j. This then fixes

$$\begin{aligned} B(E_s^{(j)}, E_{-s}^{(j)}) &= 6(-1)^{\sigma^{(j)}}\,, \\ \vec{\beta}^{(j)} &= (-1)^{\sigma^{(j)}}\vec{\alpha}^{(j)}\,, \end{aligned} \tag{3.64}$$

for some $\sigma^{(j)} = 0, 1$ to be determined.

We can now use the properties of the Killing form to determine the coefficients C_{stu}^{jkl}. From Eq. (3.54) we have that

$$B([E_s^{(j)}, E_t^{(k)}], E_u^{(l)}) = B([E_t^{(k)}, E_u^{(l)}], E_s^{(j)}) = B([E_u^{(l)}, E_s^{(j)}], E_t^{(k)})\,, \tag{3.65}$$

and using Eq. (3.17) and Eq. (3.64) we find

$$(-1)^{\sigma^{(l)}} C_{stu}^{jkl} = (-1)^{\sigma^{(j)}} C_{tus}^{klj} = (-1)^{\sigma^{(k)}} C_{ust}^{ljk}\,. \tag{3.66}$$

All the C_{stu}^{jkl} have then the same absolute value squared. Using again Eq. (3.17) we have

$$\begin{aligned} C_{stu}^{jkl} C_{-s-t-u}^{jkl} B(E_{-u}^{(l)}, E_u^{(l)}) &= B([E_s^{(j)}, E_t^{(k)}], [E_{-s}^{(j)}, E_{-t}^{(k)}]) = B(E_s^{(j)}, [E_t^{(k)}, [E_{-s}^{(j)}, E_{-t}^{(k)}]]) \\ &= B(E_s^{(j)}, [E_{-s}^{(j)}, [E_t^{(k)}, E_{-t}^{(k)}]]) = 2t(-1)^{\sigma^{(k)}} \vec{\alpha}^{(k)} \cdot B(E_s^{(j)}, [E_{-s}^{(j)}, \vec{H}]) \\ &= 2st(-1)^{\sigma^{(k)}} \vec{\alpha}^{(j)} \cdot \vec{\alpha}^{(k)} B(E_s^{(j)}, E_{-s}^{(j)})\,, \end{aligned} \tag{3.67}$$

i.e.,

$$C_{stu}^{jkl} C_{-s-t-u}^{jkl} = 2st(-1)^{\sigma^{(j)}+\sigma^{(k)}+\sigma^{(l)}} \vec{\alpha}^{(j)} \cdot \vec{\alpha}^{(k)}\,. \tag{3.68}$$

Since $C_{-s-t-u}^{jkl} = -C_{stu}^{jkl*}$, and since squaring the relation $-u\vec{\alpha}^{(l)} = s\vec{\alpha}^{(j)} + t\vec{\alpha}^{(k)}$ we obtain $1 = -2st\vec{\alpha}^{(j)} \cdot \vec{\alpha}^{(k)}$, this means

$$|C_{stu}^{jkl}|^2 = (-1)^{\sigma^{(j)}+\sigma^{(k)}+\sigma^{(l)}}\,. \tag{3.69}$$

This implies that $|C_{stu}^{jkl}| = 1$ and that $(-1)^{\sigma^{(j)}+\sigma^{(k)}+\sigma^{(l)}} = 1$, i.e., either none or two of the signs $(-1)^{\sigma^{(j)}}$ are negative. By redefining the phase of the operators $E_s^{(j)} \to e^{is\delta^{(j)}} E_s^{(j)}$, we get $C_{stu}^{jkl} \to e^{-i(s\delta^{(j)}+t\delta^{(k)}+u\delta^{(l)})} C_{stu}^{jkl}$, so that Eq. (3.66) still holds true. We can always choose to set one of the phases to a prescribed value, e.g., $C_{+--}^{2\,3\,1} = 1$, which leaves only the the relative sign of the C_{stu}^{jkl} to be determined. This is done as in the previous subsection: we multiply Eq. (3.17) on the left by $E_u^{(l)}$ to find

$$C_{stu}^{jkl} E_u^{(l)} E_{-u}^{(l)} = C_{ust}^{ljk} E_{-t}^{(k)} E_t^{(k)} + [E_s^{(j)}, E_u^{(l)} E_t^{(k)}]\,, \tag{3.70}$$

and taking the appropriate trace in the Hilbert space of the system we find that the phases must satisfy

$$C_{stu}^{jkl} = C_{ust}^{ljk} = C_{tus}^{klj}\,. \tag{3.71}$$

This then implies that

$$(-1)^{\sigma^{(j)}} = (-1)^{\sigma^{(k)}} = (-1)^{\sigma^{(l)}}\,, \tag{3.72}$$

which together with $(-1)^{3\sigma^{(j)}} = 1$ implies $\sigma^{(j)} = 0\ \forall j$. This concludes our new derivation, which leads exactly to the same results.

As we have done above, we can replace the operators $E_{\pm}^{(j)}$ with the Hermitian operators $A_1^{(j)} = \frac{1}{2}(E_+^{(j)} + E_-^{(j)})$ and $A_2^{(j)} = \frac{1}{2i}(E_+^{(j)} - E_-^{(j)})$. The algebra is then fully encoded in Eq. (3.49) with appropriate structure constants f_{abc}. It is straightforward to show that

$$\begin{aligned}
&B(A_1^{(j)}, A_1^{(k)}) = \tfrac{1}{4}B(E_+^{(j)} + E_-^{(j)}, E_+^{(j)} + E_-^{(j)}) = \tfrac{1}{2}B(E_+^{(j)}, E_-^{(j)}) = 3\,,\\
&B(A_2^{(j)}, A_2^{(k)}) = -\tfrac{1}{4}B(E_+^{(j)} - E_-^{(j)}, E_+^{(j)} - E_-^{(j)}) = \tfrac{1}{2}B(E_+^{(j)}, E_-^{(j)}) = 3\,,\\
&B(A_1^{(j)}, A_2^{(k)}) = \tfrac{1}{4i}B(E_+^{(j)} + E_-^{(j)}, E_+^{(j)} - E_-^{(j)}) = \tfrac{1}{4i}\left[B(E_+^{(j)}, E_-^{(j)}) - B(E_-^{(j)}, E_-^{(j)})\right] = 0\,,\\
&B(A_{1,2}^{(j)}, H_a) \propto B(E_+^{(j)} \pm E_-^{(j)}, H_a) = 0\,,
\end{aligned} \tag{3.73}$$

which leads to the compact expression

$$B(\mathcal{O}_a, \mathcal{O}_b) = 3\delta_{ab} \quad a, b = 1, \ldots, 8\,. \tag{3.74}$$

Let us now derive the identities of Eq. (3.50). The first one, as already stated, is obvious. The cyclicity property follows from the cyclicity of the Killing form, Eq. (3.54), together with the following result:

$$B([\mathcal{O}_a, \mathcal{O}_b], \mathcal{O}_c) = if_{abd}B(\mathcal{O}_d, \mathcal{O}_c) = 3if_{abc}\,. \tag{3.75}$$

Cyclicity and antisymmetry in the first two indices imply total antisymmetry of f_{abc}. The last identity follows from the definition of trace: since

$$\mathrm{ad}_{\mathcal{O}_a}\mathrm{ad}_{\mathcal{O}_b}\mathcal{O}_c = [\mathcal{O}_a, [\mathcal{O}_b, \mathcal{O}_c]] = -f_{bcd}f_{ade}\mathcal{O}_e\,, \tag{3.76}$$

we have that

$$3\delta_{ab} = B(\mathcal{O}_a, \mathcal{O}_b) = -f_{bcd}f_{adc} = f_{acd}f_{bcd}\,. \tag{3.77}$$

There is one more identity that we can obtain, this time by making use of the Jacobi identity:

$$\begin{aligned}
0 &= [[\mathcal{O}_a, \mathcal{O}_b], \mathcal{O}_c] + [[\mathcal{O}_c, \mathcal{O}_a], \mathcal{O}_b] + [[\mathcal{O}_b, \mathcal{O}_c], \mathcal{O}_a]\,,\\
0 &= f_{bcm}f_{amn} + f_{abm}f_{cmn} + f_{cam}f_{bmn}\,.
\end{aligned} \tag{3.78}$$

This identity can be recast as

$$-f_{bam}f_{cmn} + f_{cam}f_{bmn} = f_{bcm}f_{man}\,, \tag{3.79}$$

Notice finally that

$$\mathrm{ad}_{\mathcal{O}_a}\mathcal{O}_b = [\mathcal{O}_a, \mathcal{O}_b] = if_{abc}\mathcal{O}_c = (\mathrm{ad}_{\mathcal{O}_a})_{cb}\mathcal{O}_c \Rightarrow (\mathrm{ad}_{\mathcal{O}_a})_{bc} = if_{acb} = -if_{abc}\,. \tag{3.80}$$

These two results will prove useful later.

3.2 SU(3) and its representations

After all this fuss, do we know what symmetry group is it that we have found? Let us see if it is a matrix Lie group. If so, then there must be finite-dimensional matrices t^a satisfying the commutation relations Eq. (3.49), i.e.

$$[t^a, t^b] = i f_{abc} t^c \,. \tag{3.81}$$

These would be the matrices spanning the Lie algebra of the group. Since from the commutation relation and the identity $f_{abc}f_{abd} = 3\delta_{cd}$ we must have that $t^c = \frac{1}{3i} f_{abc}[t^a, t^b]$, these matrices must be traceless; since we need eight independent matrices, they must be at least 3×3. There are precisely eight independent, traceless, 3×3 matrices, which span the algebra $\mathfrak{su}(3)$ of the three-dimensional unitary group SU(3). Indeed, unitary matrices can be written as $U = e^{iA}$ with A Hermitian; the unimodularity condition $\det U = 1$ requires $\operatorname{tr} A = 0$. We can take the eight independent matrices as follows: $t^a = \frac{1}{2}\lambda^a$, with λ^a the *Gell-Mann matrices*

$$\begin{aligned}
\lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
\lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \qquad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\
\lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
\end{aligned} \tag{3.82}$$

The t^a obey the normalisation condition

$$\operatorname{tr} t^a t^b = \frac{1}{2}\delta^{ab} \,. \tag{3.83}$$

One can verify explicitly that the t^a satisfy Eq. (3.81). Combined with Eq. (3.83), this tells us that

$$f_{abc} = -2i \operatorname{tr} [t_a, t_b] t_c \,, \tag{3.84}$$

which shows explicitly their cyclicity property.

Given the algebra of Hermitian symmetry generators, we exponentiated it to obtain a compact group: this is needed if we want to obtain finite-dimensional unitary representations, as it was our original purpose. Our candidate for the symmetry group is then SU(3), but we still have to show that it admits the desired representations. As a matter of fact, we have already shown that the octet is indeed among the irreducible representations of this group, but this will become more clear after we briefly review representation theory.

3.2.1 Lie groups and representation theory in a nutshell

Let us briefly review Lie groups and representation theory. A *Lie group* is a group that is also a smooth manifold, with the group composition and inversion being smooth maps. A Lie group is almost entirely characterised by its *Lie algebra*, i.e., the tangent space to the identity element. If the Lie group has dimension n (the dimension of the manifold), then we can identify group

elements g in a neighbourhood of the identity by means of n real coordinates $\boldsymbol{\alpha} = \{\alpha_1 \ldots, \alpha_n\}$, i.e., $g = g(\boldsymbol{\alpha})$. The Lie algebra of the group is the linear space spanned by

$$L^a \equiv -i \frac{\partial g(\boldsymbol{\alpha})}{\partial \alpha_a}\bigg|_{\boldsymbol{\alpha}=\mathbf{0}}, \tag{3.85}$$

which are called the *generators* of the algebra or of the group, endowed with the commutator. One can show that the linear space is closed under commutators, and so one has

$$[L^a, L^b] = i C^{ab}{}_c L^c, \tag{3.86}$$

with $C^{ab}{}_c$ the *structure constants* of the group.

The proof goes as follows. One cna find a sufficiently small neighbourhood of the identity where $g(\boldsymbol{\alpha})$ $g(\boldsymbol{\beta})$ and their inverses all lie in the neighbourhood, and so does the product $g(\boldsymbol{\alpha})g(\boldsymbol{\beta})g^{-1}(\boldsymbol{\alpha})g^{-1}(\boldsymbol{\beta}) = g(\boldsymbol{\Phi}(\boldsymbol{\alpha},\boldsymbol{\beta}))$, for some function $\boldsymbol{\Phi}(\boldsymbol{\alpha},\boldsymbol{\beta})$. This function must satisfy $\boldsymbol{\Phi}(\boldsymbol{\alpha},\mathbf{0}) = \boldsymbol{\Phi}(\mathbf{0},\boldsymbol{\beta}) = 0$, since the products equal the identity element whenever one of its argument vanishes. One then finds that

$$\begin{aligned}
&\frac{\partial}{\partial\alpha_a}\frac{\partial}{\partial\beta_b} g(\boldsymbol{\Phi}(\boldsymbol{\alpha},\boldsymbol{\beta}))\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}} = \frac{\partial}{\partial\alpha_a}\left[\frac{\partial\Phi_c(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\beta_b}\frac{\partial g(\boldsymbol{\Phi}(\boldsymbol{\alpha},\boldsymbol{\beta}))}{\partial\Phi_c}\right]\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}} \\
&= \frac{\partial^2\Phi_c(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\alpha_a\partial\beta_b}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}}\frac{\partial g(\boldsymbol{\Phi})}{\partial\Phi_c}\bigg|_{\boldsymbol{\Phi}=\mathbf{0}} + \frac{\partial\Phi_c(\mathbf{0},\boldsymbol{\beta})}{\partial\beta_b}\bigg|_{\boldsymbol{\beta}=\mathbf{0}}\frac{\partial\Phi_d(\boldsymbol{\alpha},\mathbf{0})}{\partial\alpha_a}\bigg|_{\boldsymbol{\alpha}=\mathbf{0}}\frac{\partial^2 g(\boldsymbol{\Phi})}{\partial\Phi_d\partial\Phi_c}\bigg|_{\boldsymbol{\Phi}=\mathbf{0}} \\
&= \frac{\partial^2\Phi_c(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\alpha_a\partial\beta_b}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}}\frac{\partial g(\boldsymbol{\Phi})}{\partial\Phi_c}\bigg|_{\boldsymbol{\Phi}=\mathbf{0}} = i\frac{\partial^2\Phi_c(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\alpha_a\partial\beta_b}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}} L^c,
\end{aligned}$$

while on the other hand, since $\frac{\partial}{\partial\alpha_a}g^{-1} = -g^{-1}(\frac{\partial}{\partial\alpha_a}g)g^{-1}$,

$$\begin{aligned}
&\frac{\partial}{\partial\alpha_a}\frac{\partial}{\partial\beta_b} g(\boldsymbol{\Phi}(\boldsymbol{\alpha},\boldsymbol{\beta}))\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}} = \frac{\partial}{\partial\alpha_a}\left[g(\boldsymbol{\alpha})iL^b g^{-1}(\boldsymbol{\alpha})g(\mathbf{0}) - g(\boldsymbol{\alpha})g(\mathbf{0})g^{-1}(\boldsymbol{\alpha})iL^b\right]\bigg|_{\boldsymbol{\alpha}=\mathbf{0}} \\
&= \frac{\partial}{\partial\alpha_a}\left[g(\boldsymbol{\alpha})iL^b g^{-1}(\boldsymbol{\alpha}) - iL^b\right]\bigg|_{\boldsymbol{\alpha}=\mathbf{0}} = i^2(L^aL^b - L^bL^a) = -[L^a, L^b],
\end{aligned}$$

from which Eq. (3.86) follows with $C^{ab}{}_c = -\frac{\partial^2\Phi_c(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\alpha_a\partial\beta_b}\big|_{\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}}$.

From the Lie algebra, the Lie group is reconstructed via the *exponential map*,

$$g(\boldsymbol{\alpha}) = e^{i\boldsymbol{\alpha}\cdot\boldsymbol{L}}, \tag{3.87}$$

where $\boldsymbol{\alpha}\cdot\boldsymbol{L} = \sum_a \alpha_a L^a$. Actually, the same Lie algebra can be shared by different groups. What the exponential map reconstructs is the components of these groups connected to the identity, which are all isomorphic. If the group is connected and compact, then the exponential map reconstructs it entirely.

Let us now discuss the representation of groups. Given a group G (not necessarily of Lie type), a *representation* of G is a mapping D between G and the space of invertible $m \times m$ (generally complex) matrices, $GL(\mathbb{C}, n)$, that associates a matrix $D(g)$ to every element of the group, with the properties that $D(g_1g_2) = D(g_1)D(g_2)$ and $D(e) = \mathbf{1}$; this implies in particular that $D(g^{-1}) = D(g)^{-1}$. The number m is called the *dimension of the representation*. For Lie groups we consider only continuous representations, i.e., such that $D(g)$ is a continuous map from the group manifold into ${}_m\mathbb{C}_m$. Two representations D_1 and D_2 are *equivalent* if there is an

invertible mapping M such that $D_1(g) = M^{-1}D_2(g)M$ for all $g \in G$; this immediately entails that equivalent representations must have the same dimension. An *irreducible representation* is one that is not equivalent to the direct product of representations of smaller dimension, i.e., it is not possible to bring all the $D(g)$ in block diagonal form by means of a change of basis $D(g) \to M^{-1}D(g)M$; equivalently, no subspace of $\mathbb{C}^m$ is left invariant by the action of the representation.[19] For unitary groups like SU(3), and more generally for all compact Lie groups, all the irreducible representations are equivalent to unitary representations, i.e., representations in terms of unitary matrices, so that $D(U^\dagger) = D(U^{-1}) = D(U)^{-1} = D(U)^\dagger$. One can also show that for the unitary unimodular groups such matrices have to be unimodular, i.e., $\det D(U) = 1$.[20] Finally, for unitary groups the unitary representations are in bijective correspondence with the Hermitian representations of the group algebra, from which they are obtained via the exponential mapping.[21] A representation of a Lie algebra $\mathfrak{g}$ is a linear mapping d from $\mathfrak{g}$ to the space of matrices $M(\mathbb{C}, m)$, such that for all $a_1, a_2 \in \mathfrak{g}$ one has $d(a_1 + a_2) = d(a_1) + d(a_2)$ (and so $d(0) = \mathbf{0}$) and $d([a_1, a_2]) = [d(a_1), d(a_2)]$. It is irreducible if there is no subspace of $\mathbb{C}^m$ left invariant by all the $d(a)$. In practice, one finds $m \times m$ Hermitian matrices T_R^a obeying the commutation relations $[T_R^a, T_R^b] = if_{abc}T_R^c$, which provide a m-dimensional representation of the algebra, and builds the $m \times m$ unitary matrices $U_R = \exp\{i\sum_a \alpha_a T_R^a\}$, which provide a representation of the group.

3.2.2 The simplest representations of SU(3)

For any group, the simplest representation is $D(g) = 1$ for all g: this is the *trivial representation*, which is one-dimensional. Correspondingly, all the elements of the Lie algebra of the group are represented by zero.

For matrix Lie groups like SU(3), the second simplest representation is that provided by the matrix group itself. This is the *fundamental* (or *defining*) representation, $D_F(U) = U$, to which it corresponds the fundamental representation of the algebra, obtained by differentiating the group elements near the identity. For the special unitary groups SU(N), as we have already mentioned above, the corresponding Lie algebra $\mathfrak{su}(N)$ is readily found by noting that one can always write $U = e^{i\boldsymbol{\alpha}\cdot\boldsymbol{t}}$ with traceless Hermitian matrices $\boldsymbol{t} = \{t^1, \ldots, t^{N^2-1}\}$. One has that $\mathfrak{su}(N)$ is precisely the algebra of Hermitian traceless matrices, of dimension $\dim = N^2 - 1$. For SU(3), $\dim = 8$. Summarising,

$$\text{fundamental representation:} \qquad t^a \Rightarrow U = e^{i\boldsymbol{\alpha}\cdot\boldsymbol{t}}\,, \tag{3.88}$$

[19] We have been a bit imprecise here, confusing *reducible*, which means that there is an invariant subspace, with *totally reducible*, which means that can be brought in blocak-diagonal form. For semisimple groups (i.e., groups with a semisimple algebra), the two notions are equivalent.

[20] A simple argument is as follows. The subset $A \subseteq \mathrm{SU}(N)$ such that $\det[D(a)] = 1$ for $a \in A$ must be a subgroup, since $\det[D(a_1a_2)] = \det[D(a_1)D(a_2)] = \det[D(a_1)]\det[D(a_2)] = 1$, and it must be a normal subgroup, i.e., $U^\dagger aU \in A\ \forall a \in A$ and $\forall U \in \mathrm{SU}(N)$, since $\det[D(U^\dagger aU)] = \det[D(a)] = 1$. The only normal subgroups of SU(N) are the group itself and the trivial subgroup $\{\mathbf{1}\}$. In the first case, since there is at least one element for which $\det D(a) = 1$ (namely, the identity) we have that $\det[D(U)] = 1\ \forall U \in \mathrm{SU}(N)$. In the second case no other element of the group is allowed to be represented by a unimodular matrix. If we now take $U_{1,2}$ such that $[U_1, U_2] \neq 0$, which certainly exist since the group is non Abelian, and consider the element $U = U_1U_2U_1^\dagger U_2^\dagger$, we find that $\det[D(U)] = 1$, which would imply $U = \mathbf{1}$ and so $U_1U_2 = U_2U_1$, against the hypothesis, so that this second possibility is excluded.

[21] More precisely, for simply connected groups every representation of the algebra comes from a representation of the group.

where the t^a can be chosen to be $t^a = \frac{\lambda^a}{2}$ with λ^a given in Eq. (3.82), as discussed at the beginning of Section 3.2, so satisfying the commutation relations Eq. (3.81) and the normalisation condition Eq. (3.83).

If we take the complex conjugate of Eq. (3.81) we find

$$[t^a, t^b]^* = [(t^a)^*, (t^b)^*] = [(-t^a)^*, (-t^b)^*] = -if_{abc}(t^c)^* = if_{abc}(-t^c)^*\,, \tag{3.89}$$

which shows that the matrices $t^a_C \equiv -(t^a)^* = -(t^a)^T$ provide another representation for the algebra, the *complex conjugate* representation. In terms of group representations this corresponds to $e^{i\boldsymbol{\alpha}\cdot\boldsymbol{t_C}} = e^{-i\boldsymbol{\alpha}\cdot\boldsymbol{t}^*} = U^* \equiv D_C(U)$, which is indeed another valid representation since $D_C(U_1U_2) = (U_1U_2)^* = U_1^*U_2^* = D_C(U_1)D_C(U_2)$ and obviously $D_C(\mathbf{1}) = \mathbf{1}$:

$$\text{complex conjugate representation:} \qquad t^a_C = -(t^a)^* \Rightarrow U_C = U^*\,. \tag{3.90}$$

Clearly $\mathrm{tr}\, t^a_C t^b_C = \frac{1}{2}\delta^{ab}$. Both the fundamental and the conjugate representation have dimension 3, and are hence denoted as $\mathbf{3}$ and $\overline{\mathbf{3}}$, respectively. However, they are not equivalent.

There is another simple representation of the algebra which we have actually already encountered, and which is provided by the structure constants themselves. Recall Eq. (3.79), which we rewrite as

$$(-i)f_{bam}(-i)f_{cmn} - (-i)f_{cam}(-i)f_{bmn} = if_{bcm}(-i)f_{man}\,, \tag{3.91}$$

If we now set $(T^m)_{an} \equiv -if_{man}$, we find after a simple relabelling of indices

$$([T^a, T^b])_{mn} = if_{abc}(T^c)_{mn}\,. \tag{3.92}$$

The matrices T^a provide the *adjoint* representation of the algebra. From Eq. (3.80) we see that the structure constants provide the matrix elements of the operator ad_X in the basis $\{\mathcal{O}_a\}$. In other words, the adjoint representation of the algebra is provided by the action of the algebra on itself.

To see what is the corresponding representation of the group, consider the vector space spanned by the generators, i.e., the space of Hermitian traceless matrices $X = X_a t^a$, and consider the mapping

$$X \to \mathrm{Ad}_U X \equiv UXU^\dagger\,. \tag{3.93}$$

This is clearly a linear mapping of the vector space into itself, and so

$$\mathrm{Ad}_U X = X_b \mathrm{Ad}_U t^b = t^a (U_A)_{ab} X_b\,. \tag{3.94}$$

Also, Ad_U provides a representation of the group, since

$$\mathrm{Ad}_{U_1}\mathrm{Ad}_{U_2} X = U_1U_2XU_2^\dagger U_1^\dagger = (U_1U_2)X(U_1U_2)^\dagger = \mathrm{Ad}_{U_1U_2}X\,. \tag{3.95}$$

Then we can write also $(U_A)_{ab} = D_A(U)_{ab}$, and as such it can be obtained as the exponential of a representation of the algebra, $D_A(e^{i\alpha_a t^a}) = e^{i\alpha_a d_A(t^a)}$, or $U_A = e^{i\alpha_a T^a_A}$. For a U in the neighbourhood of the identity we have then

$$\begin{aligned} &t^a(U_A)_{ab}X_b = t^a(\delta_{ab} + i\epsilon_c T^c_{A\,ab} + \mathcal{O}(\epsilon^2))X_b = X + t^a i\epsilon_c T^c_{A\,ab}X_b + \mathcal{O}(\epsilon^2) \\ &= X_b U t^b U^\dagger = X_b(\mathbf{1} + i\epsilon_a t^a + \mathcal{O}(\epsilon^2))t^b(\mathbf{1} - i\epsilon_c t^c + \mathcal{O}(\epsilon^2)) = X + i\epsilon_c[t^c, t^b]X_b + \mathcal{O}(\epsilon^2) \\ &= X + t^a i\epsilon_c i f_{cba} X_b + \mathcal{O}(\epsilon^2)\,, \end{aligned} \tag{3.96}$$

i.e., $(T^c)_{A\,ab} = -if_{cab} = (T^c)_{ab}$, the generators in the adjoint representation. In a more abstract but equivalent way,

$$\begin{aligned} \mathrm{Ad}_U X &= (\mathbf{1} + i\epsilon_c(T^c) + \mathcal{O}(\epsilon^2))X = UXU^\dagger = X + i\epsilon_c[t^c, X] + \mathcal{O}(\epsilon^2) \\ &= X + i\epsilon_c \mathrm{ad}_{t^c} X + \mathcal{O}(\epsilon^2) = X + i\mathrm{ad}_{\epsilon_c t^c} X + \mathcal{O}(\epsilon^2)\,, \end{aligned} \tag{3.97}$$

i.e., the generators of the representation provided by Ad are the ad_{t^a}. This can be summarised as $\mathrm{Ad}_{e^{i\alpha_a t^a}} = e^{i\alpha_a \mathrm{ad}_{t^a}}$. Expressed as a matrix in the basis $\{t^a\}$, ad_{t^a} are provided by the structure functions, as we know from Eq. (3.80):

$$\mathrm{ad}_{t^a} X = [t^a, X] = X_b[t^a, t^b] = if_{abc}t^c X_b = t^c(-if_{acb})X_b = t^c(T^a)_{cb}X_b\,. \tag{3.98}$$

Notice that for an eigenvector of ad_{t^a} one has $\mathrm{ad}_{t^a} X = \lambda X = t^c(T^a)_{cb}X_b = t^c\lambda X_c$, i.e., $(T^a)_{cb}X_b = \lambda X_c$ if and only if $[t^a, X] = \lambda X$. This observation will come in handy below. In conclusion:

$$\text{adjoint representation:} \qquad (T^a)_{bc} = -if_{abc} \Rightarrow \begin{cases} Ut^a X_a U^\dagger = t^b (U_A)_{ba} X_a\,, \\ \mathrm{Ad}_{e^{i\alpha_a t^a}} X = e^{i\alpha_a \mathrm{ad}_{t^a}} X\,. \end{cases} \tag{3.99}$$

The dimension of this representation is equal to the number of generators, i.e., it is eight-dimensional, and for this reason it is denoted as **8**. From Eq. (3.77) one finds that

$$\mathrm{tr}\, T^a T^b = -f_{amn}f_{bnm} = f_{amn}f_{bmn} = 3\delta^{ab}\,. \tag{3.100}$$

The three representations discussed above are irreducible representations of SU(3).

It is actually a general result that given an irreducible representation T^a_R of the algebra encoded in Eq. (3.81), one finds

$$\mathrm{tr}\, T^a_R T^b_R = d_R \delta^{ab}\,, \tag{3.101}$$

with an appropriate d_R. This can be proved making use of the total antisymmetry of the structure constant (which is a consequence of the semi-semplicity and compactness of the group) as follows. Multiplying the commutation relations,

$$[T^a_R, T^b_R] = if_{abc}T^c_R\,, \tag{3.102}$$

by T^d_R and taking the trace, we find

$$\mathrm{tr}\,[T^a_R, T^b_R]T^d_R = if_{abc}\mathrm{tr}\, T^c_R T^d_R \equiv if_{abc}M_{cd}\,. \tag{3.103}$$

Using the cyclicity of the trace

$$if_{abc}M_{cd} = \mathrm{tr}\,[T^a_R, T^b_R]T^d_R = \mathrm{tr}\,[T^b_R, T^d_R]T^a_R = if_{bdc}M_{ca}\,. \tag{3.104}$$

Using antisymmetry of f_{abc} and symmetry of M_{ab} this can be recast as

$$(-if_{bac})M_{cd} = M_{ac}(-if_{bcd}) \Rightarrow T^b_A M = MT^b_A\,, \tag{3.105}$$

where T_A are the generators in the adjoint representation. We then have that $[M, T^a_A] = 0$ for all the generators of an irreducible representation of a Lie group, and by Schur's lemma this implies $M = d_R\mathbf{1}$. As we shall see below, the value of d_R is uniquely determined by the representation.[22]

[22] One could think of changing it by changing the normalisation of the generators, but that has been fixed once that the structure constants have been set to the values discussed in the previous subsection.

A direct proof of Eq. (3.101) is as follows. Given a finite-dimensional unitary representation of the algebra, since $\mathrm{tr}\,\{[A,B]C\} = \mathrm{tr}\,\{A[B,C]\}$, we find first of all that

$$s\vec{\alpha}^{\,(j)}\mathrm{tr}\,E_t^{(k)\dagger}E_s^{(j)} = \mathrm{tr}\,E_t^{(k)\dagger}[\vec{H},E_s^{(j)}] = \mathrm{tr}\,([\vec{H},E_t^{(k)}])^\dagger E_s^{(j)} = t\vec{\alpha}^{\,(k)}\mathrm{tr}\,E_t^{(k)\dagger}E_s^{(j)}\,, \tag{3.106}$$

and since $E_t^{(k)\dagger} = E_{-t}^{(k)}$,

$$(s\vec{\alpha}^{\,(j)} - t\vec{\alpha}^{\,(k)})\mathrm{tr}\,E_s^{(j)}E_{-t}^{(k)} = 0 \Longrightarrow \mathrm{tr}\,E_s^{(j)}E_{-t}^{(k)} = \delta_{st}\delta_{jk}\mathrm{tr}\,E_+^{(j)}E_-^{(j)}\,. \tag{3.107}$$

We have moreover that

$$(\vec{\alpha}^{\,(j)})^2\mathrm{tr}\,H_kE_s^{(j)} = \sum_j \mathrm{tr}\,H_k[H_j,[H_j,E_s^{(j)}]] = \sum_j \mathrm{tr}\,[H_k,H_j][H_j,E_s^{(j)}] = 0\,, \tag{3.108}$$

which implies $\mathrm{tr}\,H_kE_s^{(j)} = 0$ for all k,s,j. Finally, from Eq. (3.44) we find

$$\begin{aligned}\mathrm{tr}\,H_i[E_+^{(j)},E_-^{(j)}] &= \mathrm{tr}\,[H_i,E_+^{(j)}]E_-^{(j)} = (\vec{\alpha}^{\,(j)})_i\mathrm{tr}\,E_+^{(j)}E_-^{(j)} \\ &= \mathrm{tr}\,H_i(2\vec{\alpha}^{\,(j)}\cdot\vec{H}) = 2\mathrm{tr}\,\{H_iH_j\}(\vec{\alpha}^{\,(j)})_j\,,\end{aligned} \tag{3.109}$$

or, setting $M_{ij} = \mathrm{tr}\,\{H_iH_j\}$ and using matrix notation,

$$M\vec{\alpha}^{\,(j)} = \frac{1}{2}\mathrm{tr}\,\{E_+^{(j)}E_-^{(j)}\}\vec{\alpha}^{\,(j)}\,. \tag{3.110}$$

The vectors $\vec{\alpha}^{\,(j)}$ are then eigenvectors of M. In a two-dimensional vector space there can be at most two distinct eigenvalues; we will show now that there is actually only one eigenvalue. Applying M on both sides of the equation $\vec{\alpha}^{\,(1)} - \vec{\alpha}^{\,(2)} + \vec{\alpha}^{\,(3)} = 0$, we find

$$0 = 2M(\vec{\alpha}^{\,(1)} - \vec{\alpha}^{\,(2)} + \vec{\alpha}^{\,(3)}) = \mathrm{tr}\,\{E_+^{(1)}E_-^{(1)}\}\vec{\alpha}^{\,(1)} - \mathrm{tr}\,\{E_+^{(2)}E_-^{(2)}\}\vec{\alpha}^{\,(2)} + \mathrm{tr}\,\{E_+^{(3)}E_-^{(3)}\}\vec{\alpha}^{\,(3)}\,, \tag{3.111}$$

but this can be true only if (see footnote 17)

$$\mathrm{tr}\,\{E_+^{(1)}E_-^{(1)}\} = \mathrm{tr}\,\{E_+^{(2)}E_-^{(2)}\} = \mathrm{tr}\,\{E_+^{(3)}E_-^{(3)}\} \equiv 2d_R\,, \tag{3.112}$$

which implies that M is a multiple of the identity,

$$M_{ij} = \mathrm{tr}\,H_iH_j = d_R\delta_{ij}\,. \tag{3.113}$$

The last step is to show that

$$\begin{aligned}&\mathrm{tr}\,(E_+^{(j)} \pm E_-^{(j)})(E_+^{(j)} \pm E_-^{(j)}) = \pm 2\mathrm{tr}\,E_+^{(j)}E_-^{(j)} = \pm 4d_R\,, \\ &\mathrm{tr}\,(E_+^{(j)} \pm E_-^{(j)})(E_+^{(j)} \mp E_-^{(j)}) = \mathrm{tr}\,(E_+^{(j)}E_-^{(j)} - E_+^{(j)}E_-^{(j)}) = 0\,,\end{aligned} \tag{3.114}$$

which implies that the Hermitian operators $A_1^{(j)} = \frac{1}{2}(E_+^{(j)} + E_-^{(j)})$ and $A_2^{(j)} = \frac{1}{2i}(E_+^{(j)} - E_-^{(j)})$ satisfy

$$\mathrm{tr}\,A_m^{(j)}A_n^{(k)} = d_R\delta^{jk}\delta_{mn}\,. \tag{3.115}$$

Since $\mathrm{tr}\,H_kE_s^{(j)} = 0$, both its real and its imaginary part vanish, so that $\mathrm{tr}\,H_kA_{1,2}^{(j)} = 0$, and the statement follows.

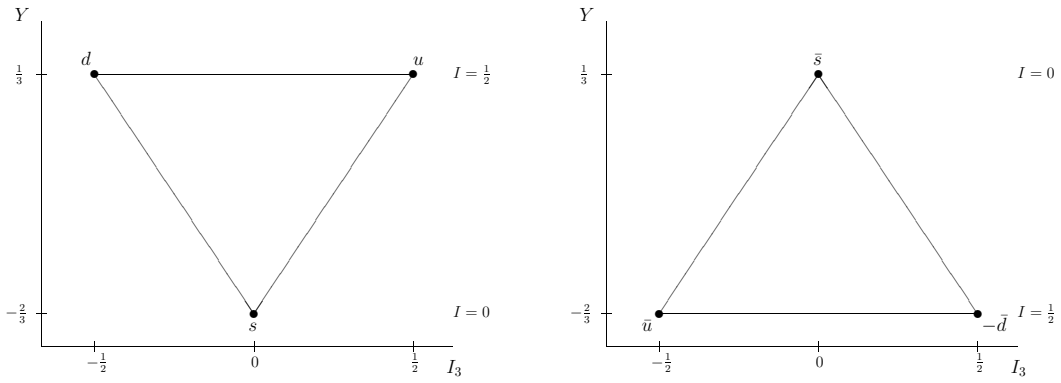

Figure 20: Weight diagrams of the fundamental and complex conjugate representations of SU(3).

3.2.3 The fundamental representation

Our discussion of representations given above does not really serve our purposes, since it is focussed on the representative matrices. What we are really concerned with are the vector spaces where the representations act, because they are the energy eigenspaces corresponding to multiplets of particles with equal mass. In fact, the representation of the group (or of its algebra) are provided by $m \times m$ matrices, which can be thought of as acting on a m-dimensional vector space, implementing the symmetry transformations corresponding to the group elements. The m basis elements correspond to m distinct particle states. A basis of this space where I_3 and Y are diagonal can potentially provide us with the physical particle states assuming of course that we have identified the correct symmetry.

An important remark is in order here. If SU(3) were truly a symmetry, any basis would be equivalent, and no choice of states would be "more physical" than any other. We know however 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 $\mathrm{SU}(2)_I$ is broken by electromagnetic interactions (as well as by small strong-interaction effects). This singles out the basis of physical particles with definite strangeness and electric charge, since only $U(1)_Y \times U(1)_{\mathrm{em}} \sim U(1)_Y \times U(1)_{I_3}$ is a true symmetry of Nature.

Let us begin with the fundamental representation of the algebra, provided by the Gell-Mann matrices via $t^a = \frac{\lambda^a}{2}$. These act on the three-dimensional space $\mathbb{C}^3$. In this representation I_3 and Y are diagonal,

$$I_3 = \frac{\lambda^3}{2} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Y = \frac{2}{\sqrt{3}}\frac{\lambda^8}{2} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \tag{3.116}$$

and so the basis vectors $e^{(j)}$, $j = 1, 2, 3$, $e_i^{(j)} = \delta_{ij}$,

$$e^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad e^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad e^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{3.117}$$

are eigenstates of I_3 and Y and can in principle represent physical states. The action of the

generators on the $e^{(j)}$ can be determined easily. Clearly,

$$\begin{aligned} I_3 e^{(1)} &= \frac{1}{2} e^{(1)}\,, \qquad & Y e^{(1)} &= \frac{1}{3} e^{(1)}\,, \\ I_3 e^{(2)} &= -\frac{1}{2} e^{(2)}\,, \qquad & Y e^{(2)} &= \frac{1}{3} e^{(2)}\,, \\ I_3 e^{(3)} &= 0 e^{(3)}\,, \qquad & Y e^{(3)} &= -\frac{2}{3} e^{(3)}\,. \end{aligned} \tag{3.118}$$

Next, we build the ladder operators,

$$\begin{aligned} I_+ &= \frac{1}{2}(\lambda^1 + i\lambda^2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,, \qquad & I_- &= \frac{1}{2}(\lambda^1 - i\lambda^2) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,, \\ V_+ &= \frac{1}{2}(\lambda^4 + i\lambda^5) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,, \qquad & V_- &= \frac{1}{2}(\lambda^4 - i\lambda^5) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}\,, \\ W_+ &= \frac{1}{2}(\lambda^6 + i\lambda^7) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\,, \qquad & W_- &= \frac{1}{2}(\lambda^6 - i\lambda^7) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,, \end{aligned} \tag{3.119}$$

from which it follows that

$$\begin{array}{llllll} I_+ e^{(1)} = 0\,, & I_- e^{(1)} = e^{(2)}\,, & V_+ e^{(1)} = 0\,, & V_- e^{(1)} = e^{(3)}\,, & W_+ e^{(1)} = 0\,, & W_- e^{(1)} = 0\,, \\ I_+ e^{(2)} = e^{(1)}\,, & I_- e^{(2)} = 0\,, & V_+ e^{(2)} = 0\,, & V_- e^{(2)} = 0\,, & W_+ e^{(2)} = 0\,, & W_- e^{(2)} = e^{(3)}\,, \\ I_+ e^{(3)} = 0\,, & I_- e^{(3)} = 0\,, & V_+ e^{(3)} = e^{(1)}\,, & V_- e^{(2)} = 0\,, & W_+ e^{(3)} = e^{(2)}\,, & W_- e^{(3)} = 0\,. \end{array} \tag{3.120}$$

Since $\vec{I}^2 = I_- I_+ + I_3(I_3+1) = I_+ I_- + I_3(I_3-1)$, we have that

$$\vec{I}^2 e^{(1)} = \frac{3}{4} e^{(1)}\,, \quad \vec{I}^2 e^{(2)} = \frac{3}{4} e^{(2)}\,, \quad \vec{I}^2 e^{(3)} = 0\,, \tag{3.121}$$

so $e^{(1)}$ and $e^{(2)}$ form an isodoublet ($I = \frac{1}{2}$) and $e^{(3)}$ is an isosinglet ($I = 0$). We can then denote the basis vectors labelling them with their eigenvalues $|I, I_3; Y\rangle$,

$$e^{(1)} = |\tfrac{1}{2}\ \tfrac{1}{2}; \tfrac{1}{3}\rangle\,, \quad e^{(2)} = |\tfrac{1}{2}\ -\tfrac{1}{2}; \tfrac{1}{3}\rangle\,, \quad e^{(3)} = |0\,0; -\tfrac{2}{3}\rangle\,. \tag{3.122}$$

Notice that all matrix elements of $I_\pm$, $V_\pm$ and $W_\pm$ are positive. The basis vectors for a representation are represented graphically as in Fig. 20, through what are called *weight diagrams*.

3.2.4 The complex conjugate representation

The complex conjugate representation also acts on $\mathbb{C}^3$, and in the basis of Eq. (3.117) is obtained by replacing $\lambda^a \to -\lambda^{a*}$. Since $-(\lambda^{3,8})^* = -\lambda^{3,8}$, we have that the eigenvalues of the basis vectors change sign,

$$\begin{aligned} I_3^C e^{(1)} &= -\frac{1}{2} e^{(1)}\,, \qquad & Y^C e^{(1)} &= -\frac{1}{3} e^{(1)}\,, \\ I_3^C e^{(2)} &= \frac{1}{2} e^{(2)}\,, \qquad & Y^C e^{(2)} &= -\frac{1}{3} e^{(2)}\,, \\ I_3^C e^{(3)} &= 0 e^{(3)}\,, \qquad & Y^C e^{(3)} &= \frac{2}{3} e^{(3)}\,. \end{aligned} \tag{3.123}$$

Moreover, since $-(\lambda^{1,4,6})^* = -\lambda^{1,4,6}$ and $-(\lambda^{2,5,7})^* = \lambda^{2,5,7}$, we have for the ladder operators that

$$I_\pm^C = -(I_\pm)^* = -I_\mp\,, \quad V_\pm^C = -(V_\pm)^* = -V_\mp\,, \quad W_\pm^C = -(W_\pm)^* = -W_\mp\,, \tag{3.124}$$

from which it follows

$$\begin{array}{llllll} I_-^C e^{(1)}{=}0\,, & I_+^C e^{(1)}{=}\text{-}e^{(2)}\,, & V_-^C e^{(1)}{=}0\,, & V_+^C e^{(1)}{=}\text{-}e^{(3)}\,, & W_-^C e^{(1)}{=}0\,, & W_+^C e^{(1)}{=}0\,, \\ I_-^C e^{(2)}{=}\text{-}e^{(1)}\,, & I_+^C e^{(2)}{=}0\,, & V_-^C e^{(2)}{=}0\,, & V_+^C e^{(2)}{=}0\,, & W_-^C e^{(2)}{=}0\,, & W_+^C e^{(2)}{=}\text{-}e^{(3)}\,, \\ I_-^C e^{(3)}{=}0\,, & I_+^C e^{(3)}{=}0\,, & V_-^C e^{(3)}{=}\text{-}e^{(1)}\,, & V_+^C e^{(2)}{=}0\,, & W_-^C e^{(3)}{=}\text{-}e^{(2)}\,, & W_+^C e^{(3)}{=}0\,. \end{array} \tag{3.125}$$

By introducing appropriate signs we can make some of the matrix elements of the ladder operators positive, but not all of them. We choose to have $I_\pm$ and $W_\pm$ with positive matrix elements only. We then set

$$e^{(1)} = |\overline{\tfrac{1}{2}\ -\tfrac{1}{2}; -\tfrac{1}{3}}\rangle\,, \quad e^{(2)} = -|\overline{\tfrac{1}{2}\ \tfrac{1}{2}; -\tfrac{1}{3}}\rangle\,, \quad e^{(3)} = |\overline{0\,0; \tfrac{2}{3}}\rangle\,, \tag{3.126}$$

where the overbar signals that this is the complex conjugate representation. Now, e.g.,

$$\begin{aligned} W_-^C|\overline{0\,0; \tfrac{2}{3}}\rangle &= W_-^C e^{(3)} = -e^{(2)} = |\overline{\tfrac{1}{2}\ \tfrac{1}{2}; -\tfrac{1}{3}}\rangle\,, \\ I_-^C|\overline{\tfrac{1}{2}\ \tfrac{1}{2}; -\tfrac{1}{3}}\rangle &= -I_-^C e^{(2)} = e^{(1)} = |\overline{\tfrac{1}{2}\ -\tfrac{1}{2}; -\tfrac{1}{3}}\rangle\,, \end{aligned} \tag{3.127}$$

while

$$V_-^C|\overline{0\,0; \tfrac{2}{3}}\rangle = V_-^C e^{(3)} = -e^{(1)} = -|\overline{\tfrac{1}{2}\ -\tfrac{1}{2}; -\tfrac{1}{3}}\rangle\,, \tag{3.128}$$

i.e., $V_\pm^C$ has a negative matrix element.[23] The weight diagram for the complex conjugate representation is also shown in Fig. 20.

So far we have found no connection between representations of SU(3) and hadronic physics. This is going to change soon.

3.2.5 The adjoint representation

To deal with the adjoint representation, let us recall that a general 3×3 Hermitian traceless matrix can be written as $X = X_a t^a$, and that

$$\begin{gathered} \mathrm{ad}(t^a)X \equiv [t^a, X] = [t^a, t^b]X_b = i f_{abc} X_b t^c = t^c (T^a)_{cb} X_b = t^c (\mathrm{ad}(t^a)X)_c \\ (\mathrm{ad}(t^a)X)_c = (T^a)_{cb} X_b\,. \end{gathered} \tag{3.129}$$

The action of the matrices T^a on $\vec{X} \in \mathbb{C}^8$, where $\vec{X}$ denotes the vector of components X_b, can then be read off the commutators $[t^a, X_m t^m]$. For example, given one of the $E_\pm^{(j)} = X(E_\pm^{(j)})_m t^m$, we have

$$\mathrm{ad}(I_3)E_\pm^{(j)} = t^c (T^3)_{cb} X (E_\pm^{(j)})_b = [I_3, E_\pm^{(j)}] = \pm\alpha_1^{(j)} E_\pm^{(j)} = \pm\alpha_1^{(j)} t^c X (E_\pm^{(j)})_c\,, \tag{3.130}$$

[23] The origin of the minus sign in Eq. (3.126) is the same discussed before when relating isospin with quarks - what a surprise...

so $\vec{X}(E^{(j)}_{\pm})$ are the eigenvectors of T^3 with eigenvalues $\pm\alpha^{(j)}_1$. It is then easy to see that for X equal to one of the ladder operators $I_\pm$, $V_\pm$, $W_\pm$ or to one of the diagonal operators I_3 and $\bar{Y} \equiv \frac{\sqrt{3}}{2}Y$ we have that

$$\begin{aligned}
T^3\vec{X}(I_3) &= 0\,, & T^8\vec{X}(I_3) &= 0\,,\\
T^3\vec{X}(\bar{Y}) &= 0\,, & T^8\vec{X}(\bar{Y}) &= 0\,,\\
T^3\vec{X}(I_\pm) &= \pm\vec{X}(I_\pm)\,, & T^8\vec{X}(I_\pm) &= 0\,,\\
T^3\vec{X}(V_\pm) &= \pm\tfrac{1}{2}\vec{X}(V_\pm)\,, & T^8\vec{X}(V_\pm) &= \pm\tfrac{\sqrt{3}}{2}\vec{X}(V_\pm)\,,\\
T^3\vec{X}(W_\pm) &= \mp\tfrac{1}{2}\vec{X}(W_\pm)\,, & T^8\vec{X}(W_\pm) &= \pm\tfrac{\sqrt{3}}{2}\vec{X}(W_\pm)\,.
\end{aligned} \tag{3.131}$$

Noticing also that

$$\mathrm{ad}(t^a)\mathrm{ad}(t^b)X = [t^a,[t^b,X]] = [t^a,t^c](T^b)_{cd}X_d = t^e(T^a)_{ec}(T^b)_{cd}X_d = t^c(T^aT^b)_{cd}X_d\,, \tag{3.132}$$

we find that

$$\begin{aligned}
\vec{I}^2_A V_\pm &= \sum_{j=1}^{3}[I_j,[I_j,V_\pm]] = \frac{3}{4}V_\pm\,, & \vec{I}^2_A W_\pm &= \sum_{j=1}^{3}[I_j,[I_j,W_\pm]] = \frac{3}{4}W_\pm\,,\\
\vec{I}^2_A I_\pm &= \sum_{j=1}^{3}[I_j,[I_j,I_\pm]] = 2I_\pm\,, & \vec{I}^2_A I_3 &= \sum_{j=1}^{3}[I_j,[I_j,I_3]] = 0\,.
\end{aligned} \tag{3.133}$$

There is now all the information we need to find out the eigenvalues and the eigenvectors of I_3 and Y in the adjoint representation. The vectors V_+, W_+, and W_-, V_- form two isodoublets ($I = \frac{1}{2}$), I_3 and $I_\pm$ form an isotriplet ($I = 1$), and $\bar{Y}$ is an isosinglet ($I = 0$). In this way the basis vectors in the adjoint representation can be chosen to be (up to phases and normalisation factors)

$$\begin{array}{ccccc}
 & |\tfrac{1}{2}\ -\tfrac{1}{2};1\rangle \propto W_+\,, & & |\tfrac{1}{2}\ \tfrac{1}{2};1\rangle \propto V_+\,, & \\
|1\ -1;0\rangle \propto I_-\,, & & |1\,0;0\rangle\,, \propto I_3 & & |1\,1;0\rangle\,, \propto I_+ \\
 & & |0\,0;0\rangle \propto Y\,, & & \\
 & |\tfrac{1}{2}\ -\tfrac{1}{2};-1\rangle \propto V_-\,, & & |\tfrac{1}{2}\ \tfrac{1}{2};-1\rangle \propto W_-\,. &
\end{array} \tag{3.134}$$

The corresponding weight diagram is shown in Fig. 21. Summarising, the algebra provides a representation of itself through the linear operators $\mathrm{ad}(t^a)$.

The important point about the adjoint representation of SU(3) is that it is precisely one of the representations we were looking for to describe hadrons, namely the one that accomodates the lightest baryons (see Fig. 18). Since there is no septuplet representation (as we will see shortly), one is led to predict the existence of an eighth light pseudoscalar meson - the η particle, with $I = Y = 0$, which was later observed experimentally.

3.2.6 General representations in a nutshell

If SU(3) is the symmetry we were looking for, then it should also explain the multiplet of baryon resonances discussed in our introductory remarks and shown in Fig. 17. In order to see if an

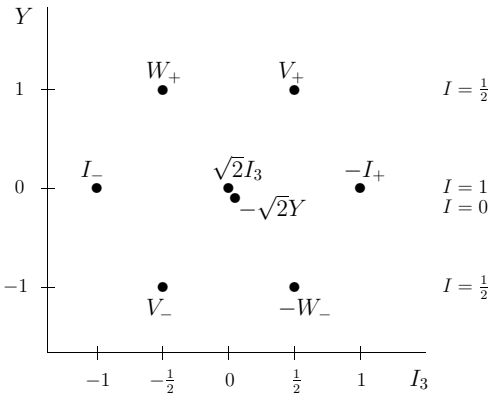

Figure 21: Adjoint representation of SU(3).

appropriate irreducible representation is available, we have to discuss representations of SU(3) in some generality. As we mentioned above, all irreducible finite-dimensional representations of SU(3) are equivalent to unitary representations, and can be reconstructed from irreducible Hermitian representations of the algebra. Let us assume that we have one such representation of the algebra, and let us figure out what it looks like.

Since I_3 and Y are Hermitian and commute, they can be diagonalised simultaneously by means of a unitary transformation, so we can take them diagonal without loss of generality (if they are not diagonal, we are going over from one unitary representation to another equivalent unitary representation). It is useful to notice that, introducing the notation $V_3 = \frac{I_3}{2} + \frac{3Y}{4} = \frac{I_3}{2} + \frac{\sqrt{3}\tilde{Y}}{2}$ and $W_3 = -\frac{I_3}{2} + \frac{3Y}{4} = -\frac{I_3}{2} + \frac{\sqrt{3}\tilde{Y}}{2}$, the three sets $\{I_j\}$, $\{V_j\}$ and $\{W_j\}$ all satisfy the SU(2) commutation relations. Taking now any eigenvector $|i_3, y\rangle$, we can build towers of eigenvectors by repeated application of the ladder operators, since $[\vec{x} \equiv (i_3, y)]$

$$\vec{H}E^{(j)}_{\pm}|i_3, y\rangle = ([\vec{H}, E^{(j)}_{\pm}] + E^{(j)}_{\pm}\vec{H})|i_3, y\rangle = (\vec{x} \pm \vec{\alpha}^{(j)})E^{(j)}_{\pm}|i_3, y\rangle\,. \tag{3.135}$$

Since the representation is finite-dimensional, the tower should end at some point, which means that after applying, say, the raising operators a certain number of times we should get zero. One can show that there must be a vector $|\psi\rangle$ such that $I_+|\psi\rangle = V_+|\psi\rangle = W_+|\psi\rangle = 0$, and that it must be unique if the representation is irreducible. This is called the *highest weight* vector.

> Existence of the highest weight can be proved as follows. Let $Z = \{|\phi^{(i)}_a\rangle\}$ be the set of all the vectors such that $I_+|\phi^{(i)}_a\rangle = 0$. This set is nonempty due to the argument above. If we now apply V_+ on these vectors, then one still has $I_+V_+|\phi^{(i)}_a\rangle = V_+I_+|\phi^{(i)}_a\rangle = 0$, i.e., all the $V^k_+|\phi^{(i)}_a\rangle$ are still in Z. Since the representation is finite-dimensional, for some value of k one must have $V^{k+1}_+|\phi^{(i)}_a\rangle = 0$. Let k_a be the smallest such value for $|\phi^{(i)}_a\rangle$, and let $V^{k_a}_+|\phi^{(i)}_a\rangle \equiv |\phi^{(iv)}_a\rangle$. Consider now repeated applyication of W_+ on the $|\phi^{(iv)}_a\rangle$. One has that $V_+W^l_+|\phi^{(iv)}_a\rangle = W^l_+V_+|\phi^{(iv)}_a\rangle = 0$, and $I_+W^l_+|\phi^{(iv)}_a\rangle = [I_+, W^l_+]|\phi^{(iv)}_a\rangle = lW^{l-1}_+V_+|\phi^{(iv)}_a\rangle = 0$. Again invoking the finite dimensionality of the representation, there must be a smallest l_a such that $W^{l_a+1}_+|\phi^{(iv)}_a\rangle = 0$. Let $W^{l_a}_+|\phi^{(iv)}_a\rangle \equiv |\phi^{(ivw)}_a\rangle$. The vectors $|\phi^{(ivw)}_a\rangle$ are such that $I_+|\phi^{(ivw)}_a\rangle = V_+|\phi^{(ivw)}_a\rangle = W_+|\phi^{(ivw)}_a\rangle = 0$.

The highest weight vector $|\psi\rangle$ is of course an eigenvector of I_3 and Y, since we took them diagonal, and so an eigenvector of V_3 and W_3, but it is an eigenvector of $\vec{I}^2$, $\vec{V}^{\,2}$ and $\vec{W}^{\,2}$ as well.

This follows from $\vec{I}^2 = I_- I_+ + I_3 + I_3^2$, and analogous formulas for $\vec{V}^{\,2}$ and $\vec{W}^{\,2}$. Let $\vec{x}_\psi = (i, y)$ be the vector of eigenvalues of $|\psi\rangle$. We can now build other states from $|\psi\rangle$ by applying the lowering operators,

$$|n_1\, n_2\, n_3\rangle = I_-^{n_1} V_-^{n_2} W_-^{n_3} |\psi\rangle\,, \tag{3.136}$$

with $n_j \geq 0$. These are all eigenvectors of $\vec{H}$ with eigenvalue $\vec{x}_{n_1\, n_2\, n_3} = \vec{x}_\psi - \sum_j n_j \vec{\alpha}^{(j)}$, as one can easily show. Due to the commutation relations, not all them are linearly independent. Using the commutation relations one can also show that applying I_+, V_+ or W_+ on any $|n_1\, n_2\, n_3\rangle$ one obtains a linear combination of the same vectors. To show that no other state besides the $|n_1\, n_2\, n_3\rangle$ can be obtained in this way one needs the following observations. Starting from $|\psi\rangle$ and applying I_- repeatedly one obtains the vectors $|n_1\, 0\, 0\rangle$, which all satisfy $V_+|n_1\, 0\, 0\rangle = W_+|n_1\, 0\, 0\rangle = 0$. These are all eigenstates of $\vec{I}^2$ with the same eigenvalue as $|\psi\rangle$. Similarly, one can apply W_- repeatedly, obtaining the vectors $|0\ 0\, n_3\rangle$, which all satisfy $I_+|0\, 0\, n_3\rangle = V_+|0\, 0\, n_3\rangle = 0$. These are all eigenstates of $\vec{W}^{\,2}$ with the same eigenvalue as $|\psi\rangle$. If we plot all the states on the (I_3, Y) plane, the states $|n_1\, 0\, 0\rangle$ and $|0\ 0\, n_3\rangle$ form a wedge which is a sort of boundary for our whole set of states. The eigenvalues corresponding to these states are non-degenerate, since there are no other states $|n_1\, n_2\, n_3\rangle$ obtained with lowering operators which can give the same eigenvalue. Next, one can show that $E_+^{(j)}|n_1\, n_2\, n_3\rangle$ must be a linear combination of those $|n_1'\, n_2'\, n_3'\rangle$ such that $\sum_k (n_k - n_k')\vec{\alpha}^{(k)} = \vec{\alpha}^{(j)}$. The action of the raising operators then allows to move between our states by following one of the root vectors $\vec{\alpha}^{(j)}$. This means that starting from any $|n_1\, n_2\, n_3\rangle$ and applying repeatedly the raising operators, at some point we will reach a state on the boundary. Due to the properties of those states, we cannot go above the upper boundary, and we cannot go right of the right boundary, no matter where we start from. The set of states $|n_1\, n_2\, n_3\rangle$ is therefore invariant under the action of the representatives of the generators of the algebra.

> If there were more than one highest weight, we could repeat this construction for each of them, and all the sets obtained in this way would be invariant under the action of the representatives of the generators. By the assumed irreducibility of the representation, this cannot happen, and so the highest weight must be unique.

Since the representation is finite-dimensional, we cannot go on and apply the lowering operators forever, and at some point their repeated action must annihilate the state. This is easy to see for the boundary states: being eigenstates of $\vec{I}^2$ and $\vec{W}^{\,2}$, they must provide representations of SU(2), which must be finite-dimensional since we are in a finite-dimensional space. Indeed, since $I_3|\psi\rangle = i|\psi\rangle$, we must have $I_-^{2i+1}|\psi\rangle = 0$: the upper boundary consists of $2i+1$ states. Similarly, since $W_3|\psi\rangle = \frac{1}{2}(-i + \frac{3}{2}y)|\psi\rangle \equiv w|\psi\rangle$, we have that $W_-^{2w+1}|\psi\rangle = 0$, and we have $2w+1$ states on the right boundary. Moreover, $2i$ and $2w$ must be zero or positive integers to have representations of SU(2), and so $2i \in \mathbb{N}_0$ and $2w = -i + \frac{3}{2}y \in \mathbb{N}_0$. Therefore $i = \frac{n}{2}$ and $y = \frac{2}{3}(i+m) = \frac{1}{3}(n+2m)$, with $n, m \in \mathbb{N}_0$. To find the remaining part of the boundary, it is useful to notice that SU(2) representations are symmetric, and are left invariant by the transformation $e^{i\pi I_2}$. This is a consequence of the well-known fact that $e^{-i\pi I_2} I_3 e^{i\pi I_2} = -I_3$. Moreover, since Y commutes with $\vec{I}$, we have $e^{-i\pi I_2} Y e^{i\pi I_2} = -I_3$. This is a particular case of the following general result: if we denote with $\vec{\alpha}_\perp^{\,(j)}$ the unit vector orthogonal to $\vec{\alpha}^{\,(j)}$, and with

$H^{(j)} = \vec{\alpha}^{\,(j)} \cdot \vec{H}$ and $H_\perp^{(j)} = \vec{\alpha}_\perp^{\,(j)} \cdot \vec{H}$, then

$$\begin{aligned} [H^{(j)}, E_s^{(j)}] &= s E_s^{(j)} \,, \\ [H^{(j)}, H_\perp^{(j)}] &= [H_\perp^{(j)}, E_s^{(j)}] = 0 \,, \\ [E_s^{(j)}, E_{-s}^{(j)}] &= 2s H^{(j)} \,. \end{aligned} \tag{3.137}$$

If we now denote with $A_2^{(j)} = \frac{1}{2i}(E_+^{(j)} - E_-^{(j)})$, this implies that

$$\begin{aligned} e^{-i\pi A_2^{(j)}} H^{(j)} e^{i\pi A_2^{(j)}} &= -H^{(j)} \,, \\ e^{-i\pi A_2^{(j)}} H_\perp^{(j)} e^{i\pi A_2^{(j)}} &= H_\perp^{(j)} \,. \end{aligned} \tag{3.138}$$

Since the representation must be invariant[24] under $e^{i\pi A_2^{(j)}}$, to each state $|\vec{x}\rangle$ with $\vec{x} = (i_3, \sqrt{3}2y)$ must correspond another state $|\vec{x}\,'\rangle$ such that

$$\begin{cases} H^{(j)}|\vec{x}\rangle = \vec{\alpha}^{\,(j)} \cdot \vec{x}|\vec{x}\rangle \\ H_\perp^{(j)}|\vec{x}\rangle = \vec{\alpha}_\perp^{\,(j)} \cdot \vec{x}|\vec{x}\rangle \end{cases} \longrightarrow \begin{cases} H^{(j)}|\vec{x}\,'\rangle = -\vec{\alpha}^{\,(j)} \cdot \vec{x}|\vec{x}\,'\rangle \\ H_\perp^{(j)}|\vec{x}\,'\rangle = \vec{\alpha}_\perp^{\,(j)} \cdot \vec{x}\,'|\vec{x}\rangle \end{cases} \tag{3.139}$$

The set of possible eigenvalues, and thus our plot of the states $|n_1\, n_2\, n_3\rangle$ in the $(I_3, \bar{Y})$ plane, must then be invariant under the reflections

$$\vec{x} \to \vec{x} - 2(\vec{x} \cdot \vec{\alpha}^{\,(j)})\vec{\alpha}^{\,(j)} \,. \tag{3.140}$$

This applies in particular to $\vec{\alpha}^{\,(1)}$ and $\vec{\alpha}^{\,(3)}$. The composition of two reflections is a rotation: reflecting first around $\vec{\alpha}_\perp^{\,(1)}$ and then $\vec{\alpha}_\perp^{\,(3)}$ results in $\vec{\alpha}^{\,(1)} \to -\vec{\alpha}^{\,(2)}$, which is a clockwise rotation of 60°. If the plot in the $(I_3, \bar{Y})$ plane is symmetric under these transformations, so must be its boundary, which is then determined to be that of an irregular hexagon (which can degenerate into a triangle), with upper side of length $2i$, and with sides of alternating length $2i$ and $2w$. This completely fixes the shape of our representation. A possible weight diagram is shown in Fig. 22.

These results can be used, together with uniqueness of the highest weight, to show that any state can be connected to any other state by the action of the representatives of the algebra. To do so, it suffices to connect any state to the highest weight by means of raising operators. If for some state this could not be done, then repeated application of V_+ must annihilate the state before the boundary is reached, for otherwise one could use I_+ or W_+ to move from there and reach $|\psi\rangle$. Let $|\phi\rangle$ be the state obtained in this way such that $V_+|\phi\rangle = 0$. Repeated application of I_+ on $|\phi\rangle$ yields states that are still annihilated by V_+, and before reaching the boundary on must find $I_+^{k+1}|\phi\rangle = I_+|\phi'\rangle = 0$. Repeated application of W_+ must now yield $W_+^{l+1}|\phi'\rangle = W_+|\phi''\rangle = 0$ before the boundary is reached. Since $V_+|\phi''\rangle = W_+^l V_+|\phi'\rangle = 0$, and $I_+|\phi''\rangle = [I_+, W_+^l]|\phi'\rangle = l W_+^{l-1} V_+|\phi'\rangle = 0$, we have that $|\phi''\rangle$ would be another highest weight, contraddicting irreducibility.

What remains to be done is identify the independent states among the $|n_1\, n_2\, n_3\rangle$ and fix the degeneracy of each point in the weight diagram. One can show that the degeneracy increase by

[24] This does not mean that each state is left invariant, but it means that the set of states as a whole is left invariant.

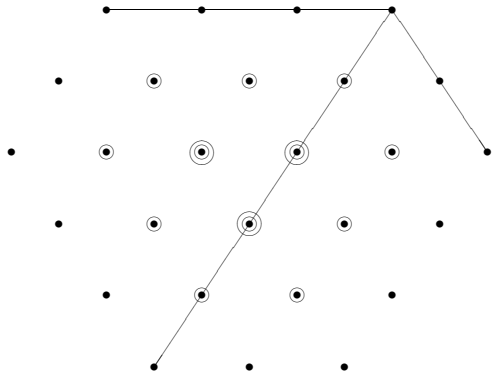

Figure 22: Weight diagram for a possible representation of SU(3).

one for each further hexagonal layer, until the first triangular layer is reached: from that point on the degeneracy does not change anymore. One can further show that states with a given value of Y can be organised in multiplets with definite $\vec{I}^2$: starting from the upper boundary, which is an $I = i$ multiplet with $2i+1$ states and $Y = y$, one the second line $Y = y-1$ one finds an $I = i+\frac{1}{2}$ multiplet and a $I = i-\frac{1}{2}$ multiplet, for a total of $(2i+1+1)+(2i+1-1) = 2(2i+1)$ states (as it also results counting from the layers). In general, on line $l+1$, where $Y = y-l$, one finds one isospin multiplet for each of the values of I between $i_{\min}$ and $i_{\max}$, where:

- $i_{\max} = i + \frac{l}{2}$, as long as $l \leq 2w$: this is where the top right part of the boundary ends; beyond this point, $i_{\max} = i + 2w - \frac{l}{2}$;
- $i_{\min} = i - \frac{l}{2}$, as long as $l \leq 2i$: this is where $i_{\min}$ hits zero; beyond this point, $i_{\min} = \frac{l}{2} - i$;
- when $i_{\min} = i_{\max}$ the construction ends: this happens when $l = 2(i+w)$.

Notice that since the highest weight is also an eigenvector of $\vec{V}^2$ with eigenvalue v, where $2v = i + \frac{3}{2}y$, by descending from it via V_- one must hit zero when applying it $2v+1$ times, so there cannot be more than $2v+1$ lines. But $2(i+w) = 2i - i + \frac{3}{2}y = 2v$, so there are indeed $2v+1 = 2(i+w)+1$ lines.

3.2.7 Obtaining the baryonic multiplets: the "eightfold way"

Let us apply this machinery to obtain again the spin-$\frac{1}{2}$ baryon octet, and in an attempt to reproduce the pattern of the spin-$\frac{3}{2}$ baryon resonances. For the octet, the particle in the top right corner, the proton, has $I = I_3 = \frac{1}{2}$ and $Y = 1$, so $i = w = \frac{1}{2}$. From this state we obtain the neutron, $I = -I_3 = \frac{1}{2}$ and $Y = 1$, by applying I_-. No further states with $Y = 1$ are available. If we now apply W_- we obtain a state with $I_3 = 1$ which also has automatically $I = 1$, and moreover $Y = 0$. Applying I_- repeatedly we obtain the isospin triplet of the Sigmas, Σ^+, Σ^0, Σ^-, all with $Y = 0$ (i.e., $S = -1$). Applying instead the combination of lowering operators $\frac{1}{2}V_- - I_-W_-$ on the proton, we obtain a state with $Y = 0$, $I_3 = 0$ and such that $I_+(\frac{1}{2}V_- - I_-W_-)|p\rangle = 0$, which is then an eigenstate of total isospin with $I = 0$: this is the Λ. Finally, applying V_- on the Σ^+ we obtain a state with $Y = -1$ and $I_3 = \frac{1}{2}$, which is also automatically an eigenstate of $\vec{I}^2$: indeed, since $I_+|\Sigma^+\rangle = W_-|\Sigma^+\rangle = 0$, we have $I_+V_-|\Sigma^+\rangle = [I_+, V_-]|\Sigma^+\rangle = W_-|\Sigma^+\rangle = 0$. This state is then the $I_3 = \frac{1}{2}$ component of a doublet ($I = \frac{1}{2}$) of particles with strangeness $S = -2$,

i.e., the Ξ^0; the $I_3 = -\frac{1}{2}$ component of the doublet, the Ξ^-, is obtained by applying I_-. No further states can be built.

The construction above was just an exercise, since we knew already of the existence of the octet (i.e., the adjoint) representation. Let us now see if we can accommodate the baryon resonances. The particle in the top-right corner is the Δ^{++}, which has $i = \frac{3}{2}$ and $y = 1$, and which we tentatively take as our highest weight $|\psi\rangle$. We then have $w = 0$, so the hexagon degenerates into an equilateral triangle of side length $2i = 3$. Since $W_-|\psi\rangle = 0$, one has automatically that $I_+V_-|\psi\rangle = 0$ (see above), and so $V_-|\psi\rangle$ is an eigenstate of $\vec{I}^2$, with $I = i - \frac{1}{2} = 1$. The same applies to the states obtained through successive applications of V_- starting from $|\psi\rangle$: there are in total three such states, each one an eigenstate of $\vec{I}^2$ with $I = 1$. $\frac{1}{2}$, and 0. From these states one obtains isospin multiplets applying I_-; each multiplet also has a definite value of Y. All in all there are the four states with $I = \frac{3}{2}$ and $Y = 1$ obtained from $|\psi\rangle$ via I_-; the three states with $I = 1$ and $Y = 0$ obtained from $V_-|\psi\rangle$ via I_-; the two states with $I = \frac{1}{2}$ and $Y = -1$ obtained from $V_-^2|\psi\rangle$ via I_-; and one state with $I = 0$ and $Y = -2$ given by $V_-^3|\psi\rangle$. No more states can be constructed: using $W_- = [V_-, I_+]$ on any of these states gives a linear combination of the others.

In group-thoretical terms, this shows that SU(3) possesses a ten-dimensional representation – the decuplet. Nine of the states in the decuplet perfectly matched the then-known resonances, so if the symmetry explanation was true, a tenth particle had to exist, with $I = 0$, strangeness $S = Y - 1 = -3$ and electric charge $Q = I_3 + \frac{1}{2}Y = -1$. Let us look at the mass pattern of the resonances: $m_\Delta = 1232$ MeV, $m_{\Sigma^*} = 1384$ MeV, $m_{\Xi^*} = 1533$ MeV, so that $m_{\Sigma^*} - m_\Delta = 152$ MeV and $m_{\Xi^*} - m_{\Sigma^*} = 149$ MeV. An educated guess, estimating $m(S) = m(\Delta) + 150\ \text{MeV} \cdot |S|$ would place the misterious tenth particle, let us call it Ω^-, at a mass of around $m_\Omega = 1682$ MeV.[25] Notice that a similar mass formula approximately works for the octet as well. In 1964 a spin-$\frac{3}{2}$ baryon resonance with the predicted properties was indeed observed at a mass $m_\Omega = 1672$ MeV.

The classification of hadron multiplets in terms of irreducible representations of SU(3) goes under the rather bizarre name of *eightfold way*, and was proposed by Murray Gell-Mann in 1961. A similar proposal was made independently by Y. Ne'eman, also in 1961. It is also worth remembering here the model of composite hadrons of S. Sakata and related developments, which played an important role in suggesting SU(3) as the relevant approximate symmetry.

3.3 From SU(3) 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 could not explain why only certain irreducible representation appeared in Nature, while other did not. From the group-theoretical point of view all representations are good, but Nature showed only baryon octets (like the lightest spin-$\frac{1}{2}$ baryons) and decuplet (like the spin-$\frac{3}{2}$ baryonic resonances), and meson octets (like the lightest pseudoscalar mesons K, π and η) and singlets (like the pseudoscalar meson η'). Quite surprisingly, the fundamental and complex conjugate representation do not appear.

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 (**3**)

[25] The estimate was done actually through a more accurate formula, the Gell-Mann–Okubo formula, discussed below.

and complex conjugate ($\overline{\mathbf{3}}$) representations.[26] In particular, the simplest ways to produce an octet are via the composition $\mathbf{3} \otimes \overline{\mathbf{3}}$ of a fundamental and a conjugate, or via the composition $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3}$ of three fundamentals. Interestingly enough, the first composition produces an octet and a singlet, $\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$, while the second one produces a decuplet, two octets and a singlet, $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}$. The derivation of these results is discussed below, while here we would rather focus on what they suggest. The composition of a fundamental and a complex conjugate representations yields precisely the representations appearing in the classification of mesons, while the composition of three fundamental representations yields precisely the representations appearing in the classification of baryons, plus an unobserved singlet representation. Together with results such as the Gell-Mann–Nishijima formula $Q = I_3 + \frac{1}{2}Y = I_3 + \frac{1}{2}(B + S)$, and the approximate mass relation $m(S) = m(0) + 150 \text{ MeV} \cdot |S|$ valid within the octet and decuplet baryon multiplets (and other results as well), this "naturally" suggests that hadrons are not elementary, but composite particles,[27] built out of three constituents that form a basis for a fundamental representation of SU(3), and their antiparticles. The elementary constituents became known as *quarks*, another bizarre name courtesy of M. Gell-Mann, who used them 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.

If one assumes that quarks are the elementary constituents of hadrons, one can hope to solve the representation puzzle using the constraints coming from the exchange symmetry of identical particles. Before trying that, though, one should be able to assign to the quarks appropriate values of the various quantum numbers that can reproduce the observed phenomenology, besides the values of I_3 and Y that come from them being a basis for the fundamental representation. We will denote the three eigenstates of I_3 and Y with u, d and s, with eigenvalues assigned according to Fig. 20, or Tab. 10. The type of quark is called the quark *flavour*. The corresponding representatives in the space $\mathbb{R}^3$ where the representation acts are the vectors $e^{(i)}$, $i = 1, 2, 3$ discussed in Section 3.2.3.

Let us now discuss the other quantum numbers. First of all, since we know that baryon number must commute with all the SU(3) generators, we can assign a baryon number to quarks, and the same number to each of them. Since three quarks are required to form a baryon, we automatically find $B_u = B_d = B_s = \frac{1}{3}$. Next come electric charge and strangeness. As we will see in a moment, composing representations and reducing them to their irreducible components, one can find what is the quark content of each baryon. In particular, one finds 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 to determine the value of charge and strangeness of each quark:

$$\begin{array}{lcccc} p\text{:} & 2Q_u + Q_d = 1\,, & & 2S_u + S_d = 0\,, \\ n\text{:} & Q_u + 2Q_d = 0\,, & & S_u + 2S_d = 0\,, \\ \Lambda\text{:} & Q_u + Q_d + Q_s = 0\,, & & S_u + S_d + S_s = -1\,, \end{array} \tag{3.141}$$

[26] Actually the fundamental representation suffices, as we will see below.

[27] This all look natural from a modern perspective, but when 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.

	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 10: Quantum number assignement of quarks.

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.

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 charges, like I_3, Y, and baryon number. More precisely, QFT demands that we associate to them the same representative vector $e^{(i)}$ as the corresponding particle, so that for each flavour charge conjugation is implemented as $Cf = \bar{f}$,[28], and that we impose that they transform according to the complex-conjugate representation of SU(3). 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 are $s = \frac{1}{2}$ fermions, quarks should be assigned spin $s = \frac{1}{2}$, so that one can obtain $\frac{1}{2}$ out of the composition $\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \frac{1}{2} \oplus \frac{1}{2} \oplus \frac{3}{2}$.[29] 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.

Now that we are done assigning quantum numbers, it is time to work out explicitly the composition of SU(3) representations, which we have already used above. The easiest way is to do it graphically, taking into account that the tensor product of two states will have eigenvalues of I_3 and Y equal to the sum of the eigenvalues of those two states. This is depicted in Fig. 23 for the tensor product of two fundamental representations: we just draw in the (I_3, Y)-plane a triangle corresponding to the first **3**, and three more triangles centered on the vertices of the first one, reaching out to the eigenvalues obtained by adding those of two eigenvectors of the fundamental representation. We then decompose the result into irreducible representations: here we are guided by symmetry considerations in assigning states in degenerate (I_3, Y)-eigenspaces to different representations. We then get from the figure that $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \bar{\mathbf{3}}$. A similar procedure shows that $\mathbf{6} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8}$, and that $\bar{\mathbf{3}} \otimes \mathbf{3} = \mathbf{8} \oplus \mathbf{1}$. Putting everything together

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = (\mathbf{6} \oplus \bar{\mathbf{3}}) \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}\,. \tag{3.142}$$

Irrespectively of the precise form of the linear combinations of products of three quarks 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. Denoting with $n_{u,d,s}$ the

[28] We could introduce a phase 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?

[29] The minimal spin that can be obtained from three spin-$\frac{n}{2}$ particles is just $\frac{n}{2}$: assuming naturally that for the lightest states the relative orbital angular momenta vanish, one is forced to take $n = 1$.

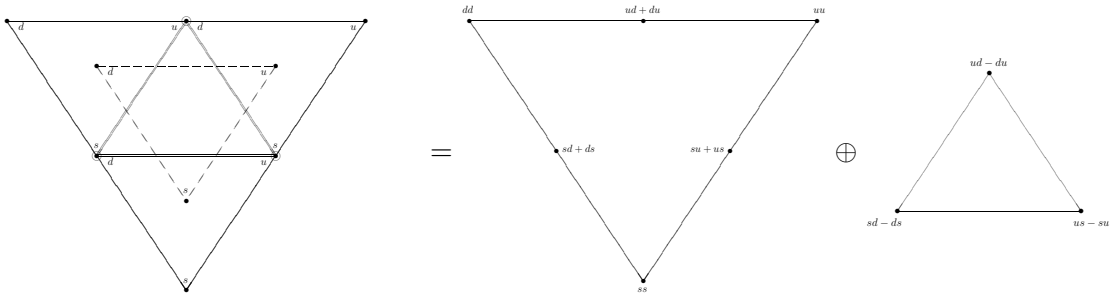

Figure 23: Composing representations of SU(3): $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} + \bar{\mathbf{3}}$.

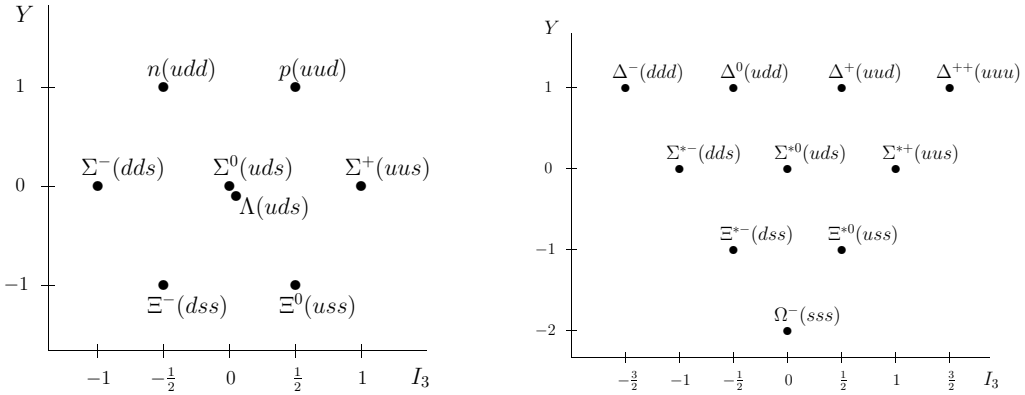

Figure 24: Quark content of octet (left) and decuplet (right) baryons.

number of times each flavour appears, since of course $n_u + n_d + n_s = 3B$ with $B = 1$, and

$$I_3 = \tfrac{1}{2}(n_u - n_d)\,, \qquad Y = \tfrac{1}{3}(n_u + n_d - 2n_s)\,, \tag{3.143}$$

which can be inverted to give

$$n_u = I_3 + \tfrac{1}{2}Y + B = I_3 + \tfrac{1}{2}Y + 1\,, \quad n_u = -I_3 + \tfrac{1}{2}Y + B - I_3 + \tfrac{1}{2}Y + 1\,, \quad n_s = B - Y = 1 - Y\,. \tag{3.144}$$

This result implies the quark composition of the proton, the neutron and the Λ used above. Moreover, using the relation $\bar{\mathbf{3}} \otimes \mathbf{3} = \mathbf{8} \oplus \mathbf{1}$ we can construct the meson multiplets, and we can similarly obtain the quark/antiquark composition of each meson, counting each antiquark as minus one quark, and solving Eq. (3.143) with the constraint $n_u + n_d + n_s = 0$. This yields

$$n_u = I_3 + \tfrac{1}{2}Y\,, \qquad n_u = -I_3 + \tfrac{1}{2}Y\,, \qquad n_s = -Y\,. \tag{3.145}$$

The quark content of octet and decuplet baryons is shown in Fig. 24, while that of octet and singlet mesons is shown in Fig. 25.

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

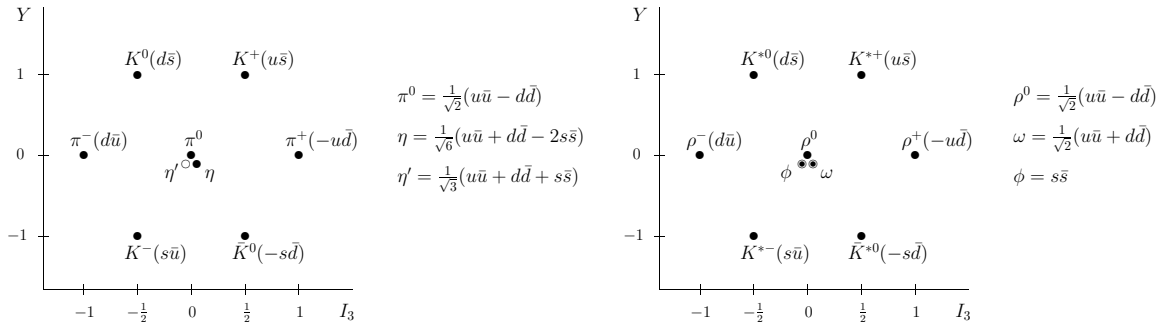

Figure 25: Quark content of octet and singlet pseudoscalar (left) and vector (right) mesons. The singlet pseudoscalar is denoted with an empty dot; the physical states coming from the mixing of the $I = 0$ octet and of the singlet states in the vector case are denoted with a half-filled dot.

$m_B = n_u m_u + n_d m_d + n_s m_s$. Within an isospin multiplet $\Delta m_B = m_1 - m_2 = (n_{u1} - n_{u2})m_u + (n_{d1} - n_{d2})m_d = (n_{u1} - n_{u2})(m_u - m_d)$, but since these splittings are known to be very small, we set in a first approximation $m_u = m_d$, and thus $m_B = m_u(n_u + n_d) + m_s n_s = 3m_u + (m_s - m_u)|S|$. From the mass splittings we then find $m_s - m_u = 150$ MeV, while from the mass of the nucleon we get $3m_u \approx 940$ MeV. We then find $m_u \simeq m_d \approx 300$ MeV, $m_s \approx 450$ 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 comes from the fact that actually, contrary to our assumption, most of hadron masses 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.

What we have discussed above are the basics of the *quark model* (Gell-Mann, 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, and to explain baryon masses it is required to introduce an explicit breaking of the symmetry, through a different assignement of masses to the various flavours. Since we now know what the microscopic theory of strong interactions is (namely, QCD), we know exactly what kind of symmetry breaking term will appear. The form of this term was however guessed correctly before the discovery of QCD, and led to the Gell-Mann–Okubo formula to be discussed below. Before doing that, however, we have to fix a serious problem of the quark model.

3.3.1 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 statistics, and this can hopefully explain why only certain representations appear. In fact, 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 $\psi = \psi_{\text{space}}\psi_{\text{spin}}\psi_{\text{flavour}}$, 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.

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:

$$\begin{aligned}
\Delta^{++} &= |uuu\rangle\,, \\
\Delta^{+} &\propto I_- \Delta^{++} \propto |uud\rangle + |udu\rangle + |duu\rangle\,, \\
\Sigma^{*+} &\propto V_- \Delta^{++} \propto |uus\rangle + |usu\rangle + |suu\rangle\,, \\
\Sigma^{*0} &\propto V_-^2 \Delta^{++} \propto |uss\rangle + |sus\rangle + |ssu\rangle\,, \\
\Omega^{-} &\propto V_-^3 \Delta^{++} \propto |sss\rangle\,,
\end{aligned} \tag{3.146}$$

and similarly for the other states. 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 a fully antisymmetric singlet,

$$\frac{1}{6}\left(|uds\rangle + |dsu\rangle + |sud\rangle - |usd\rangle - |sdu\rangle - |dus\rangle\right)\,, \tag{3.147}$$

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 first two quarks yield the antisymmetric state

$$\frac{1}{\sqrt{2}}\left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\right)\frac{1}{\sqrt{2}}\left(|ud\rangle - |du\rangle\right)\,. \tag{3.148}$$

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. (3.148) involving quarks i and j, and by ϕ_k the flavour wave function of the k-th quarks, and similarly denoting with $\tilde{\psi}_{ij}$ the antisymmetric spin wave function of Eq. (3.148) involving quarks i and j, and by $\tilde{\phi}_k$ the spin wave function of the k-th quarks, we can build

$$\begin{aligned}
&\psi_{12}\phi_3\tilde{\psi}_{12}\tilde{\phi}_3 + \psi_{23}\phi_1\tilde{\psi}_{23}\tilde{\phi}_1 + \psi_{13}\phi_2\tilde{\psi}_{13}\tilde{\phi}_2 \\
&= |udd;\uparrow\downarrow\uparrow\rangle + |dud;\downarrow\uparrow\uparrow\rangle + |ddu;\uparrow\downarrow\uparrow\rangle + |udd;\uparrow\uparrow\downarrow\rangle + |dud;\uparrow\uparrow\downarrow\rangle + |ddu;\downarrow\uparrow\uparrow\rangle \\
&\quad - 2|udd;\downarrow\uparrow\uparrow\rangle - 2|dud;\uparrow\downarrow\uparrow\rangle - 2|ddu;\uparrow\uparrow\downarrow\rangle\,,
\end{aligned} \tag{3.149}$$

which is totally symmetric. Note that $\psi_{12} + \psi_{23} = \psi_{13}$: the three functions are not independent, and they should not 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 shuld be the case since we have only to 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 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 pseudscalar 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. 25.

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 it in general it should also be a singlet of $\mathrm{SU}(N_c)$, for all hadrons. For N_c colours, the wave function is then either $\epsilon_{i_1 \ldots i_{N_c}}$ or $\delta_{i_1 i_2}$.[30] The wave function $\delta_{i_1 i_2}$ corresponds to a fundamental/complex conjugate pair, and so is appropriate for mesons. The Levi-Civita tensor is instead totally antisymmetric, and if one assumes that there are $N_c = 3$ colours one achieves two results at once: it provides an "explanation" of why it takes three quarks to make a baryon, and one can combine it 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 (Gell-Mann, Leutwyler, Fritsch, 1972). Besides using theoretical considerations to fix it, the number of colours can also be determined experimentally by studying 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. The 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. Since the probability of annihilating into a photon is the same for any quark colour, the probability of this process is proportional to N_c; experiments show that $N_c = 3$. Another experimental confirmation comes from the neutral pion decay process $\pi^0 \to \gamma\gamma$, which is also proportional to N_c for the same reasons.

3.3.2 The Gell-Mann–Okubo formula

As 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 $m_s \gg m_u, m_d$, while

[30] More complicated wave functions are also possible, but let us stick to the simplest ones.

$m_u \simeq m_d$, we can write down the strong interaction Hamiltonian in the quark rest frame as

$$\begin{aligned}\langle q_i|H|q_j\rangle = M_{ij} = m_i\delta_{ij} = \mathrm{diag}(m_{ud}, m_{ud}, m_s) \\ = \frac{2m_{ud}+m_s}{3}\mathbf{1} + \frac{m_{ud}-m_s}{3}\mathrm{diag}(1,1,-2) = \frac{2m_{ud}+m_s}{3}\mathbf{1} + \frac{m_{ud}-m_s}{\sqrt{3}}\lambda_8\,.\end{aligned} \tag{3.150}$$

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.[31] 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). 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. Given an exactly degenerate multiplet of baryons in the SU(3) symmetric case, with $H_0|B^{(0)}\rangle = m^{(0)}|B^{(0)}\rangle$, what we have to do is to diagonalise the perturbation within the degenerate subspace, i.e., diagonalise $\langle B^{(0)\prime}|H_8|B^{(0)}\rangle$. The ground-state baryons consist of an octet and a decuplet, so they cannot be mixed by the perturbation, and we need worry only about the matrix elements of H_8 between members of a single irreducible representation. 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.

In order to see this, we will need a few pieces of information about representations, and about SU(3). First of all consider a generic matrix element $\langle B'(R')|H_8|B(R)\rangle$ between baryon states B and B' transforming in the representations R and R', respectively. This object transforms according to the representation $R \otimes \bar{R}'$, since

$$\langle B'(R')|\hat{U}^\dagger H_8 \hat{U}|B(R)\rangle = \sum_{\tilde{B},\tilde{B}'} (U_{R'})^*_{\tilde{B}'B'}(U_R)_{\tilde{B}B}\langle \tilde{B}'(R')|H_8|\tilde{B}(R)\rangle\,, \tag{3.151}$$

where $\hat{U}$ is the unitary operator effecting flavour rotations on the Hilbert space of states. It is a general fact that the tensor product $R\otimes\bar{R}'$ can be decomposed into irreducible representations, $R \otimes \bar{R}' = \bigoplus_{\tilde{R}} \tilde{R}$, with a given irreducible representation appearing possibly more than once. This means that we can write

$$\langle B'(R')|H_8|B(R)\rangle = \sum_{\tilde{R},i} C_i^{(\tilde{R})}(R,R';H_8)(\mathcal{T}_i^{(\tilde{R})})_{B'B}\,, \tag{3.152}$$

where the $\mathcal{T}_i^{(\tilde{R})}$ transform as

$$\sum_{\tilde{B},\tilde{B}'} (U_{R'})^*_{\tilde{B}'B'}(U_R)_{\tilde{B}B}(\mathcal{T}_i^{(\tilde{R})})_{\tilde{B}'\tilde{B}} = \sum_j (U_{\tilde{R}})_{ji}(\mathcal{T}_j^{(\tilde{R})})_{B'B}\,, \tag{3.153}$$

[31] This actually remains true in the full dynamical case of QCD, where the symmetry-breaking term is a mass term proprtional 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 $\int d^3x\psi^\dagger\lambda_8\psi$.

and the coefficients $C_i^{(\tilde{R})}$ depend on the representations R and R', and on the details of the interaction. Asking for H_8 to transform like the eighth member of an octet amounts to set all these coefficients to zero, except, those corresponding to $\tilde{R} = \mathbf{8}$ and $i = 8$. As we said above, this can happen more than once, but for SU(3) and for $R' = R$ this cannot happen more than twice. This means that the most general form of the matrix elements we are interested in reads

$$\langle B'(R)|H_8|B(R)\rangle = \delta m_1(R)(\mathcal{T}_8^{(\mathbf{8},1)})_{B'B} + \delta m_2(R)(\mathcal{T}_8^{(\mathbf{8},2)})_{B'B}\,, \tag{3.154}$$

for two independent tensorial structures $\mathcal{T}_8^{(\mathbf{8},j)}$, and two unknown, multiplet-dependent coefficients $\delta m_j(R)$. The first structure can obviously be taken to be $\mathcal{T}_8^{(\mathbf{8},1)} = T_R^8$, since one has $U_R^\dagger T_R^a U_R = (U_{\mathbf{8}})_{ab} T_R^b$ independently of the representation.

To find the second structure we have to go back to the Gell-Mann matrices, and notice that the product of any two of them is necessarily of the form

$$\lambda^a \lambda^b = \frac{2}{3}\delta_{ab} + (if_{abc} + d_{abc})\lambda^c\,. \tag{3.155}$$

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 Hermtian 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 is fixed by the normalisation of the Gell-Mann matrices, the f_{abc} are just the structure constants of the group, and the symbols

$$d_{abc} = \tfrac{1}{4}\mathrm{tr}\,\{\lambda^a,\lambda^b\}t^c = 2\mathrm{tr}\,\{t^a,t^b\}t^c \tag{3.156}$$

are totally symmetric since they are invariant under cyclic permutations of the indices. An important property is that d_{abc} is invariant under the transformation of all the indices via the adjoint representation, $(U_{\mathbf{8}})_{a'a}(U_{\mathbf{8}})_{b'b}(U_{\mathbf{8}})_{c'c}d_{a'b'c'} = d_{abc}$, which entails that $D_R^a \equiv d_{abc}T_R^bT_R^c$ transforms in the adjoint, $U_R^\dagger D_R^a U_R = U_R^\dagger d_{abc}T_R^bT_R^cU_R = (U_{\mathbf{8}})_{aa'}d_{a'bc}T_R^bT_R^c = (U_{\mathbf{8}})_{aa'}D_R^{a'}$. We then have

$$\langle B'(R)|H_8|B(R)\rangle = \delta m_1(R)(T_R^8)_{B'B} + \delta m_2(R)(D_R^8)_{B'B}\,, \tag{3.157}$$

The values of d_{8bc} are explicitly known, and using them one finds

$$\begin{aligned} D_R^8 = d_{8bc}T_R^bT_R^c &= -\frac{1}{2\sqrt{3}}\sum_a (T_R^a)^2 + \frac{\sqrt{3}}{2}[(T_R^1)^2 + (T_R^2)^2 + (T_R^3)^2] - \frac{1}{2\sqrt{3}}(T_R^8)^2 \\ &= -\frac{1}{2\sqrt{3}}C_R + \frac{\sqrt{3}}{2}\left(\vec{I}^2 - \frac{1}{4}Y^2\right)\,. \end{aligned} \tag{3.158}$$

Here $C_R = \sum_a (T_R^a)^2$ is the quadratic *Casimir operator*, which commutes with all the generators and therefore with all the elements of the irreducible group representation. By Schur's lemma, it must be proportional to the identity within the multiplet. We then find that the perturbation is diagonal within a multiplet, and the diagonal terms read

$$\Delta m_B = \langle B^{(0)}|H_8|B^{(0)}\rangle = \frac{\sqrt{3}}{2}\left[\delta m_1 Y + \delta m_2\left(-\frac{C_R}{3} + I(I+1) - \frac{1}{4}Y^2\right)\right]\,. \tag{3.159}$$

Putting everything together and redefining appropriately the unknown constants, we obtain

$$m_B = m^{(0)} + \langle B^{(0)}|H_8|B^{(0)}\rangle = \tilde{m}^{(0)} + \delta\tilde{m}_1 Y + \delta\tilde{m}_2 \left[I\,(I+1) - \frac{1}{4}Y^2 \right] . \tag{3.160}$$

It is understood that the coefficients depend on the irreducible multiplet. This is the *Gell-Mann–Okubo mass formula*. Let us check it in practice.

Baryon octet For the baryon octet we have (the notation $X_{I,Y}$ is used for particle X):

$$\begin{aligned}
&\Lambda_{0,0}: && m_\Lambda = \tilde{m}^{(0)} \\
&N_{\frac{1}{2},1}: && m_N = \tilde{m}^{(0)} + \delta\tilde{m}_1 + \frac{1}{2}\delta\tilde{m}_2 \\
&\Sigma_{1,0}: && m_\Sigma = \tilde{m}^{(0)} + 2\delta\tilde{m}_2 \\
&\Xi_{\frac{1}{2},-1}: && m_\Xi = \tilde{m}^{(0)} - \delta\tilde{m}_1 + \frac{1}{2}\delta\tilde{m}_2 \,,
\end{aligned} \tag{3.161}$$

Since there are four equations with three unknowns, one can exctract one relation among masses. This can be taken to be for example

$$m_N + m_\Xi = \tfrac{3}{2} m_\Lambda + \tfrac{1}{2} m_\Sigma \,. \tag{3.162}$$

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

$$I(I+1) - \frac{1}{4}Y^2 = 2 + \frac{3}{2}Y \,, \tag{3.163}$$

and so the mass formula boils down to

$$m_B = \tilde{m}^{(0)} + \delta m Y \,. \tag{3.164}$$

This explains the very accurate linear dependence of decuplet masses on strangeness. In this case one finds

$$\begin{aligned}
&\Delta_{\frac{3}{2},1}: && m_\Delta = \tilde{m}^{(0)} + \delta m \,, \\
&\Sigma^*_{1,0}: && m_{\Sigma^*} = \tilde{m}^{(0)} \,, \\
&\Xi^*_{\frac{1}{2},-1}: && m_{\Xi^*} = \tilde{m}^{(0)} - \delta m \,, \\
&\Omega_{0,-2}: && m_\Omega = \tilde{m}^{(0)} - 2\delta m \,,
\end{aligned} \tag{3.165}$$

from which one can extract two mass relations, e.g.,

$$\begin{aligned}
m_\Delta + m_{\Xi^*} &= 2m_{\Sigma^*} \,, \\
2m_\Delta + m_\Omega &= 3m_{\Sigma^*} \,,
\end{aligned} \tag{3.166}$$

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 Omega was initially predicted by means of this type of 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. (3.162),

$$4m_K^2 = 3m_\eta^2 + m_\pi^2 \tag{3.167}$$

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.

4 Weak interactions

As we have discussed in the Introduction, besides the symmetry-rich strong interactions there are the wildly symmetry-breaking weak ones, mediated by the intermediate vector bosons $W^\pm$ and Z^0. The weak interactions mix the quark flavours through the CKM matrix, so that quark flavour is no longer a conserved quantity: in particular, strangeness is no more conserved. Nevertheless, those particular flavour combinations corresponding to electric charge and baryon number are still exactly conserved. Also isospin is no longer a symmetry.

An important consequence of this is that a few particles that, due to symmetry reasons, are stable against decay when only electromagnetic and strong interactions are taken into account, can now decay via weak interactions. For example, in the meson pseudoscalar octet only the π^0 was unstable against decay (e.g. $\pi^0 \to \gamma\gamma$ was possible): charged pions could not decay because of u and d flavour conservation since there is no lighter state with the same flavour content; the decay into electron plus photon, allowed by phase space considerations only, is forbidden by lepton number conservation. Similarly, kaons cannot decay due to conservation of strangeness, since they are the lightest strange particles. With weak interactions at play, pions can decay since flavour is no more a good quantum number and since there are neutrinos available to conserve lepton number; kaons can decay since strangeness is no more conserved.

Other symmetries violated by weak interactions are parity P and charge conjugation C, as first demonstrated by Wu's experiment in 1956 (after Yang and Lee's proposal). In this experiment, polarised nuclei of cobalt 60, ${}^{60}_{27}\text{Co}$ ($J = 5$) undergo β-decay into an excited state of nichel 60, ${}^{60}_{28}\text{Ni}^*$ ($J = 4$), which promptly decays electromagnetically to its ground state emitting two photons:

$$\begin{array}{rcl} {}^{60}_{27}\text{Co} & \longrightarrow & {}^{60}_{28}\text{Ni}^* + e^- + \bar{\nu}_e\,, \\ & & \quad \hookrightarrow {}^{60}_{28}\text{Ni} + \gamma + \gamma \end{array} \tag{4.1}$$

where the fundamental process taking place is the β-decay of the neutron, $n \to p\, e^-\, \bar{\nu}_e$. Cooling down the cobalt nuclei and putting them in a uniform magnetic field one can make their spin point in the, say, up direction. Since the electron and the antineutrino are both spin-$\frac{1}{2}$, angular momentum conservation requires that the nichel nucleus, the electron and the antineutrino are all polarised in the up direction. Since the nichel nucleus is essentially produced at rest, the electron and the antineutrino are emitted back-to-back to conserve momentum. If parity were a symmetry, since spin does not change sign under it while momentum does, it would be equally

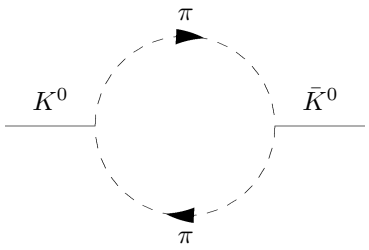

Figure 26: K^0–$\bar{K}^0$ oscillations through a pion loop.

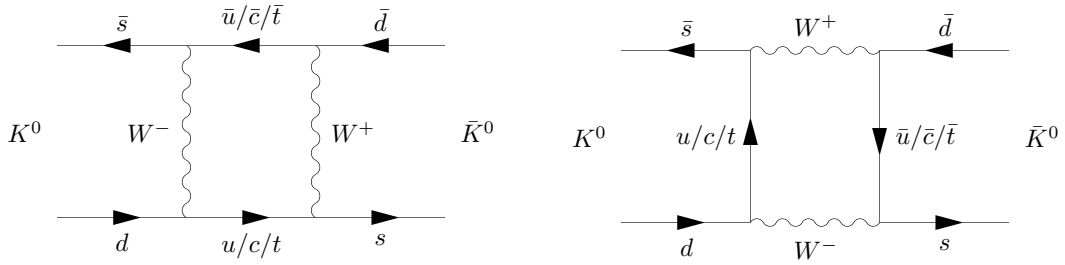

Figure 27: K^0–$\bar{K}^0$ oscillations in the modern view.

probable to find electrons emitted in the direction of the nuclear spin and in the direction opposite to it. Experiment shows that this is not the case, and that electrons are preferentially emitted in the direction opposite to the nuclear spin. This inequivocably shows violation of parity. The explanation of this result comes from the fact that antineutrinos only exist with positive helicity h, where $h = \frac{\vec{p}\cdot\vec{s}}{|\vec{p}|}$. If the electron is emitted exactly in the direction opposite to that of the nuclear spin, then $h_e = -1$ and $h_{\bar{\nu}} = +1$, which is all right; if the electron were emitted exactly in the direction of the nuclear spin, then its helicity would be $h_e = +1$, and that of the antineutrino would be $h_{\bar{\nu}} = -1$, which is impossible.

If we were to study instead the decay of anticobalt into antinichel, we would produce a positron and an electronic neutrino, which has negative helicity: the same argument as above shows that positrons would be emitted preferentially in the direction of the nuclear spin instead of the opposite one: this implies that charge conjugation symmetry is violated as well. However, if we combined P and C and observed the positron in the parity-reflected world, then we would again see it being emitted in the direction opposite to the nuclear spin: we might then hope that the combined CP operation is still a symmetry even in the presence of weak interactions. For reasons that we will now explain, the optimal place to look for CP violations is the K^0, $\bar{K}^0$ system (and the B^0, $\bar{B}^0$ system as well, where the B-mesons are made of down and bottom (anti)quarks).

The peculiarity of the two neutral K mesons is that they have the same quantum numbers except for strangeness, and so if strangeness is not anymore a conserved quantity there is nothing preventing them from mixing. Indeed, even before the modern theory of weak interactions was formulated, the fact that both the K^0 and the $\bar{K}^0$ can decay into two pions implies that they can oscillate into each other through a virtual pion loop (see Fig. 26). In modern terms, the oscillation takes place through a second-order process via W-exchange (see Fig. 27). The existence of oscillations in the neutral kaon system implies that the weak Hamiltonian has non-vanishing matrix elements between the K^0 and the $\bar{K}^0$ states. Notice that these are eigenstates

of the strong interactions, through which they are produced in the laboratory. Their decay, on the other hand, proceeds through the weak interactions.

If the kaons could mix but were otherwise stable, the unitary temporal evolution of a kaon state $|\psi(t)\rangle$ would be limited to the kaon subspace,

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle = c_1(t)|K^0\rangle + c_2(t)|\bar{K}^0\rangle\,, \tag{4.2}$$

and it would therefore be unitary within this subspace. Since kaon decays, the evolution leads the state outside this subspace, i.e.,

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle = c_1(t)|K^0\rangle + c_2(t)|\bar{K}^0\rangle + |R(t)\rangle\,, \tag{4.3}$$

with $|R(t)\rangle$ orthogonal to the kaon states. If we look at the projection of the state at time t on the kaon subspace and compare it to the state at time 0, which we assume to be in this subspace, then we would not observe a unitary evolution. Formally, denoting with Π_K the projector on the kaon subspace,

$$|\psi_K(t)\rangle \equiv \Pi_K|\psi(t)\rangle = \Pi_K U(t)|\psi(0)\rangle = \Pi_K U(t)\Pi_K|\psi(0)\rangle = c_1(t)|K^0\rangle + c_2(t)|\bar{K}^0\rangle\,, \tag{4.4}$$

and $\Pi_K U(t)\Pi_K = \Pi_K e^{-iHt}\Pi_K$ is not unitary, where $H = H_{\text{strong}} + H_{\text{weak}}$. However, under certain approximations one can show that $\Pi_K e^{-iHt}\Pi_K \simeq e^{-iH_{\text{eff}}t}$ with an effective non-Hermitian Hamiltonian $H_{\text{eff}} \neq H_{\text{eff}}^\dagger$: this is the *Weisskopf-Wigner approximation*. The non-Hermiticity precisely reflects the fact that kaons can decay. The temporal evolution is then governed by the equation

$$i\frac{\partial}{\partial t}|\psi_K(t)\rangle = H_{\text{eff}}|\psi_K(t)\rangle\,, \tag{4.5}$$

which is most easily solved by finding eigenvalues and eigenvectors of the effective Hamiltonian. The eigenvalues are in general complex, and the eigenvectors non-orthogonal. Let us denote the eigenvectors with $|K_S\rangle$ and $|K_L\rangle$, subject to the normalisation condition $\langle K_S|K_S\rangle = \langle K_L|K_L\rangle = 1$, and the corresponding eigenvalues with $\lambda_{S,L} = m_{S,L} - i\frac{\Gamma_{S,L}}{2}$, with real $m_{S,L}$ and $\Gamma_{S,L}$. We take by convention $\Gamma_S \geq \Gamma_L$. If we take $|\psi(0)\rangle = |\psi_K(0)\rangle = c_S|K_S\rangle + c_L|K_L\rangle$, then

$$|\psi_K(t)\rangle = c_S e^{-i\left(m_S - i\frac{\Gamma_S}{2}\right)t}|K_S\rangle + c_L e^{-i\left(m_L - i\frac{\Gamma_L}{2}\right)t}|K_L\rangle\,. \tag{4.6}$$

The real part of the eigenvalues is easily interpreted as the mass of these states, while the imaginary part is seen to correspond to their decay width. In fact, the norm of the projection of the state on the kaon subspace is given at time t by the expression

$$|\langle\psi_K(t)|\psi_K(t)\rangle|^2 = |c_S|^2 e^{-\Gamma_S t} + |c_L|^2 e^{-\Gamma_L t} + 2\text{Re}\left\{c_S^* c_L e^{i(m_S - m_L)t} e^{-\frac{\Gamma_S+\Gamma_L}{2}t}\langle K_S|K_L\rangle\right\}\,, \tag{4.7}$$

from which the interpretation of $\Gamma_{S,L}$ follows. This shows that in general, as far as the decay process goes, there are two types of kaons, $|K_S\rangle$ and $|K_L\rangle$, with different masses $m_{S,L}$ and decay times $\Gamma_{S,L}^{-1}$, the $|K_S\rangle$ being shorter-lived than the $|K_L\rangle$. For this reason, the $|K_S\rangle$ is named "K-short", while the $|K_L\rangle$ is named "K-long". Moreover, in general neither K_S nor K_L coincide with the strong interaction eigenstates K^0 and $\bar{K}^0$, so if the initial state is one of these, it will have both c_S and c_L nonzero, and from Eq. (4.6) we see that at some later times it will develop a component corresponding to the other kaon.

So far we have discussed oscillation and decay of the neutral mesons in general terms. Let us now suppose that CP is a symmetry of the weak interactions. This implies that CP commutes both with the projector Π_K (since it is a symmetry of strong interactions, and the kaon subspace is a CP eigenspace) and with the full Hamiltonian H. But then, in the approximation we are working in,

$$0 = [CP, \Pi_K e^{-iHt} \Pi_K] = [CP, e^{-iH_{\text{eff}}t}] , \tag{4.8}$$

i.e., $[CP, H_{\text{eff}}] = 0$, and so the two operators can be diagonalised simultaneously. This implies that $CP|K_{S,L}\rangle = \eta_{S,L}|K_{S,L}\rangle$, with some phases $\eta_{S,L}$. Since K^0 and $\bar{K}^0$ are pseudoscalars, if we choose phases so that $C|K^0\rangle = |\bar{K}^0\rangle$ (recall that QFT requires $C^2 = 1$) then the states

$$|K_1\rangle = \frac{1}{\sqrt{2}} \left(|K^0\rangle - |\bar{K}^0\rangle \right) , \qquad |K_2\rangle = \frac{1}{\sqrt{2}} \left(|K^0\rangle + |\bar{K}^0\rangle \right) , \tag{4.9}$$

are eigenstates of CP with $CP|K_1\rangle = |K_1\rangle$ and $CP|K_2\rangle = -|K_2\rangle$. Then $K_{S,L}$ can be identified with $K_{1,2}$, and so $\eta_{S,L} = \pm 1$, and $\langle K_S | K_L \rangle = 0$. Before 1956 it was known that kaons decayed into two pions in $\tau_S \simeq 0.89 \cdot 10^{-10} s$. Since both kaons and pions are spinless, the final state must have $\ell = 0$ and so positive parity[32] $(-1)^\ell \eta_\pi^2 = 1$, and so positive CP phase as well, $CP = 1$ (notice that pions have $C = 1$). A $CP = -1$ state, on the other hand, cannot decay into two pions. This lead to identify the $CP = 1$ state K_1 with the K_S. A $CP = -1$ state can decay instead into three pions, since $(-1)^0(-1)^3 = -1$, so we are led to identify K_2 with the K_L. In 1956 Lederman and collaborators observed kaons decaying into three pions further down the beam: this was the "long" component of the kaon, with $\tau_L = 5.2 \cdot 10^{-8} s$.

Let us summarise the situation so far. Kaons are produced via strong interactions as eigenstates of strangeness K^0 and $\bar{K}^0$, that can be expressed in terms of eigenstates of CP as

$$|K^0\rangle = \frac{1}{\sqrt{2}} \left(|K_1\rangle + |K_2\rangle \right) , \qquad |\bar{K}^0\rangle = -\frac{1}{\sqrt{2}} \left(|K_1\rangle - |K_2\rangle \right) . \tag{4.10}$$

In the decay process, which is governed by weak interactions, strangeness is not conserved anymore; if CP is still a symmetry, then K_1 and K_2 are the states with a definite lifetime, and a beam of kaons will see its K_1 component ($CP = 1$), identified with K_S, to decay first (mostly into two pions),[33] and its K_2 component ($CP = -1$), identified with K_L, to decay later (mostly into three pions, never in two pions). We note in passing that the mass difference between "long" and "short" kaons is really tiny, $m_L - m_S \simeq 3.5 \cdot 10^{-6}$ eV.

What if CP were not a symmetry of weak interactions? In this case there would be no reason for the physical state K_L to be a pure $CP = -1$ state, i.e., to be pure K_2, and in general it will contain a K_1 component as well,

$$|K_L\rangle = \frac{1}{\sqrt{1 + |\varepsilon|^2}} \left(|K^2\rangle + \varepsilon |K_1\rangle \right) , \tag{4.11}$$

and down the beam where K_S has already decayed we would observe some CP-forbidden decays in two pions. In 1964 Cronin, Fitch and collaborators, using a very long kaon beam, observed

[32] Notice that parity is different in the initial and final state.

[33] Notice that K_1 decaying into a three pion state $\pi^+\pi^-\pi^0$ is possible if angular momentum is chosen properly in the final state. If we take $\ell = 1$ in the $\pi^+\pi^0$ pair, and $\ell = 1$ for the π^0 with respect to the center of mass of the $\pi^+\pi^-$ pair, then we can have $J = 0$ overall. Under CP, $CP|((\pi^+\pi^-)_{\ell=1}\pi^0)_{\ell=1}\rangle = -|((\pi^-\pi^+)_{\ell=1}\pi^0)_{\ell=1}\rangle = -(-1)^1|((\pi^+\pi^-)_{\ell=1}\pi^0)_{\ell=1}\rangle$, so $CP = 1$.

such two-pion decays of the K-long: this was enough to show that CP does not commute with H_{eff}, and so is violated by weak interactions. This is a rather small violation: experimentally $|\varepsilon| = 2.2 \cdot 10^{-3}$. This manifestation of CP non-invariance is called *indirect* CP violation, and it is due to the physical states not being CP eigenstates. It is also possible to have *direct* CP violation: since CP is not anymore a symmetry, even a $CP = -1$ eigenstate like K_2 can decay in two pions, as the symmetry is not there anymore to forbid this process. Such direct violations of CP have been actually observed experimentally.

We note in passing that in order to have explicit CP breaking in the Standard Model it is necessary that the CKM matrix be at least 3×3, thus requiring the existence of at least three families of quarks: in this way a CP-violating phase remains in the matrix after all the irrelevant phases have been reabsorbed in a redefinition of fermion states. Kobayashi and Maskawa then suggested the existence of a third family even before the existence of the charm quark, completing the second one, had been taken seriously.

5 Scattering theory

The most important type of experiment in particle physics is the scattering experiment: a bunch of particles is thrown against a fixed target, or against another bunch of particles, and what comes out of the collision is carefully analysed. 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.[34] 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. 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 project the state of the system on some particle state with definite particle content and particle momenta.

5.1 Formal theory of scattering

Let us discuss a bit more in detail what it means that the state of the system looks like a freely-evolving particle state. In mathematical 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 $e^{-iH_0t}|\phi_i$ and $e^{-iH_0t}|\phi_f$, respectively. This means

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

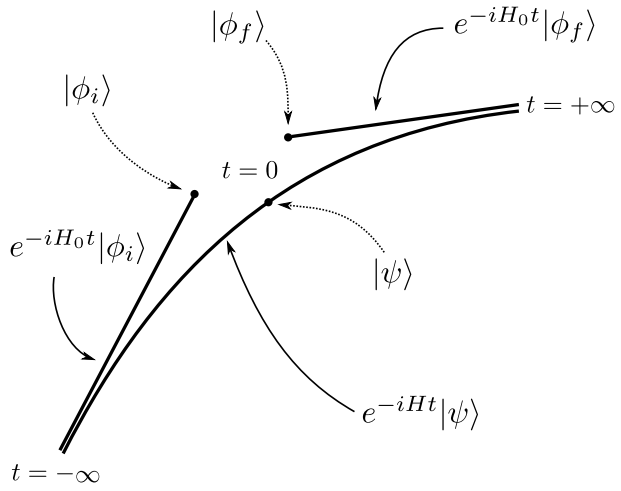

Figure 28: Schematic depiction of the relation between free and exact evolution of the system in a scattering process.

that for states $|\psi\rangle$ describing a scattering system one has

$$\begin{aligned} \lim_{t\to-\infty} \|e^{-iHt}|\psi\rangle - e^{-iH_0t}|\phi_i\rangle\| = 0\,, \\ \lim_{t\to+\infty} \|e^{-iHt}|\psi\rangle - e^{-iH_0t}|\phi_f\rangle\| = 0\,, \end{aligned} \tag{5.1}$$

for certain $|\phi_{i,f}\rangle$. A pictorial representation of the situation is given in Fig. 28. 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

$$|\psi_+\rangle = \lim_{t\to-\infty} e^{iHt}e^{-iH_0t}|\phi_i\rangle\,. \tag{5.2}$$

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[35]

$$|\psi_-\rangle = \lim_{t\to+\infty} e^{iHt}e^{-iH_0t}|\phi_f\rangle\,. \tag{5.3}$$

The states $|\psi_+\rangle$ and ψ_- 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. Eqs. (5.2) and (5.3) define two operators, the *scattering* (or *Møller*) operators,

$$\Omega_\pm = \lim_{t\to\mp\infty} e^{iHt}e^{-iH_0t}\,. \tag{5.4}$$

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 (when

[35] 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 to 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 $t \to +\infty$.

it is free). From this 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.[36] If so, then for all states $|\psi\rangle$ there are asymptotic initial and final states, i.e., the limits $\lim_{t\to\mp\infty} e^{iH_0 t} e^{-iHt}$ exist. These are nothing but the operators $\Omega_\pm^\dagger$, and then since $\forall|\psi_\pm\rangle\ \exists|\phi_\pm$ such that $|\psi_\pm\rangle = \Omega_\pm|\phi_\pm\rangle$, we have

$$|\phi_\pm\rangle = \Omega_\pm^\dagger\Omega_\pm|\psi_\pm\rangle = \Omega_\pm^\dagger|\psi_\pm\rangle \Rightarrow \Omega_\pm\Omega_\pm^\dagger|\psi_\pm\rangle = \Omega_\pm|\phi_\pm\rangle = |\psi_\pm\rangle\,, \tag{5.5}$$

for the full complete set of states $|\psi\rangle$. From this we find that $\Omega_\pm\Omega_\pm^\dagger = \mathbf{1}$, and therefore that $\Omega_\pm$ are unitary operators.

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 $|\psi_+(t)\rangle = e^{-iHt}|\psi_+\rangle \to e^{-iH_0 t}|\phi_i\rangle$ 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 $|\psi_-(t)\rangle \to e^{-iH_0 t}|\phi_f\rangle$ as $t \to +\infty$, the relevant transition amplitude is given by

$$\langle\psi_-(T_f)|\psi(T_f)\rangle = \langle\psi_-|e^{iHT_f}e^{-iHT_f}|\psi_+\rangle = \langle\psi_-|\psi_+\rangle\,, \tag{5.6}$$

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

$$\begin{aligned}
\langle\psi_-|\psi_+\rangle\,, &= \lim_{T_f\to+\infty}\langle\psi_-|e^{iHT_f}e^{-iHT_f}|\psi_+\rangle \\
&= \lim_{T_f\to+\infty,T_i\to-\infty}\langle\psi_-|e^{iHT_f}e^{-iHT_f}e^{iHT_i}e^{-iHT_i}|\psi_+\rangle \\
&= \lim_{T_f\to+\infty,T_i\to-\infty}\langle\phi_f|e^{iH_0T_f}e^{-iHT_f}e^{iHT_i}e^{-iH_0T_i}|\phi_i\rangle \\
&= \langle\phi_f|\Omega_-^\dagger\Omega_+|\phi_i\rangle \equiv \langle\phi_f|S|\phi_i\rangle = S_{fi}
\end{aligned} \tag{5.7}$$

where we have defined the *S-operator*, $S = \Omega_-^\dagger\Omega_+$, whose matrix elements S_{fi} constitute the *S-matrix*. The operator S is unitary, being the product of unitary operators, so $S^\dagger S = SS^\dagger = \mathbf{1}$. 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.

Let us discuss a few important properties of S. First of all, notice that $\forall s$

$$\begin{aligned}
e^{iHs}\Omega_\pm e^{-iH_0 s} &= \lim_{t\to\mp\infty} e^{iHs}e^{iHt}e^{-iH_0 t}e^{-iH_0 s} = \lim_{t\to\mp\infty} e^{iH(t+s)}e^{-iH_0(t+s)} \\
&= \lim_{t\to\mp\infty} e^{iHt}e^{-iH_0 t} = \Omega_\pm\,.
\end{aligned} \tag{5.8}$$

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

$$H\Omega_\pm = \Omega_\pm H_0\,, \qquad H_0\Omega_\pm^\dagger = \Omega_\pm^\dagger H\,. \tag{5.9}$$

[36] 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 it does not constitute a problem.

It then follows that

$$H_0 S = H_0 \Omega_-^\dagger \Omega_+ = \Omega_-^\dagger H \Omega_+ = \Omega_-^\dagger \Omega_+ H_0 \Rightarrow [H_0, S] = 0\,, \tag{5.10}$$

i.e., energy is conserved in a scattering process – as it should. Next, 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 chosen as to be translationally and rotationally invariant, so for the momentum $\vec{P}$ and the angular momentum $\vec{J}$ we have

$$[\vec{P}, \Omega_\pm] = [\vec{J}, \Omega_\pm] = 0\,. \tag{5.11}$$

From this it follows immediately that

$$[\vec{P}, S] = [\vec{J}, S] = 0\,, \tag{5.12}$$

i.e., momentum and angular momentum are conserved in scattering processes – again, as it should be.

Let us now work out a useful formula for the S operator. Starting from its definition, we can write

$$S = \Omega_-^\dagger \Omega_+ \lim_{t_2 \to +\infty} \lim_{t_1 \to -\infty} e^{iH_0 t_2} e^{-iH t_2} e^{iH t_1} e^{-iH_0 t_1} = \lim_{t_2 \to +\infty} \lim_{t_1 \to -\infty} \mathcal{U}(t_2, t_1)\,, \tag{5.13}$$

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 obey, and solve it subject to the initial condition $\mathcal{U}(t, t) = \mathbf{1}$. Taking the derivative with respect to t_2

$$\frac{\partial}{\partial t_2} \mathcal{U}(t_2, t_1) = e^{iH_0 t_2} i(H_0 - H) e^{-iH_0 t_2} \mathcal{U}(t_2, t_1) = -i e^{iH_0 t_2} V e^{-iH_0 t_2} \mathcal{U}(t_2, t_1) = -i V_I(t_2) \mathcal{U}(t_2, t_1)\,, \tag{5.14}$$

where

$$V_I(t) \equiv e^{iH_0 t} V e^{-iH_0 t}\,. \tag{5.15}$$

Notice that the temporal evolution is governed by the free Hamiltonian: for this reason, the subscript I is used to distinguish $V_I(t)$ from the operator $V(t) = e^{iHt} V e^{-iHt}$ in the Heisenberg picture. More on this later. 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:

$$\frac{\partial}{\partial t_1} \mathcal{U}(t_2, t_1) = \mathcal{U}(t_2, t_1) i V_I(t_1)\,. \tag{5.16}$$

The solution of Eqs. (5.14) and (5.16) with the prescribed initial condition is

$$\begin{aligned} \mathcal{U}(t_2, t_1) &= \mathrm{Texp}\left\{ -i \int_{t_1}^{t_2} dt\, V_I(t) \right\} \\ &= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int_{t_1}^{t_2} d\tau_1 \ldots \int_{t_1}^{t_2} d\tau_n T\left\{ V_I(\tau_1) \ldots V_I(\tau_n) \right\}\,, \end{aligned} \tag{5.17}$$

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

$$\begin{aligned} T\{A_1(t_1)A_2(t_2)\} &= \theta(t_1-t_2)A_1(t_1)A_2(t_2) + \theta(t_2-t_1)A_2(t_2)A_1(t_1)\,, \\ T\{A_1(t_1)\dots A_n(t_n)\} &= \sum_P \theta(t_{P(1)}-t_{P(2)})\dots\theta(t_{P(n-1)}-t_{P(n)})A_{P(1)}(t_{P(1)})\dots A_{P(n)}(t_{P(n)})\,, \end{aligned} \tag{5.18}$$

where the sum is over all the distinct permutations P of $\{1,\dots,n\}$. To see that the expression in Eq. (5.17) is indeed the solution of our equation, we rewrite it identically as follows,

$$\mathcal{U}(t_2,t_1) = \sum_{n=0}^{\infty}(-i)^n \int_{t_1}^{t_2} d\tau_1 \int_{t_1}^{\tau_1} d\tau_2 \dots \int_{t_1}^{\tau_{n-1}} d\tau_n\, V_I(\tau_1)\dots V_I(\tau_n)\,, \tag{5.19}$$

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

$$\begin{aligned} \frac{\partial}{\partial t_2}\mathcal{U}(t_2,t_1) &= -iV_I(t_2)\sum_{n=1}^{\infty}(-i)^{n-1}\int_{t_1}^{t_2} d\tau_2 \dots \int_{t_1}^{\tau_{n-1}} d\tau_n\, V_I(\tau_2)\dots V_I(\tau_n) \\ &= -iV_I(t_2)\sum_{n=0}^{\infty}(-i)^{n}\int_{t_1}^{\tau_1} d\tau_1 \dots \int_{t_1}^{\tau_{n-1}} d\tau_n\, V_I(\tau_1)\dots V_I(\tau_n) \\ &= -iV_I(t_2)\mathcal{U}(t_2,t_1)\,. \end{aligned} \tag{5.20}$$

Since it is obvious that it satisfies the initial condition, it is the unique solution of our problem. In terms of $\mathcal{U}$ we can then write *Dyson's formula* for the S-operator:

$$S = \mathcal{U}(+\infty,-\infty) = \mathrm{Texp}\left\{-i\int_{-\infty}^{+\infty} dt\, V_I(t)\right\}. \tag{5.21}$$

As promised, we briefly comment on the I in $V_I(t)$. As is well known, quantum mechanics can be equivalently formulated in the Schrödinger picture, in which states evolve with time as determined by the (full) Hamiltonian of the system, while observables are independent of time,

$$\text{Schrödinger:}\quad |\psi(t)\rangle_S = e^{-iHt}|\psi(0)\rangle_S\,, \qquad \mathcal{O}_S\,, \tag{5.22}$$

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

$$\text{Heisenberg:}\quad |\psi\rangle_H = |\psi(0)\rangle_S\,, \qquad \mathcal{O}_H(t) = e^{iHt}\mathcal{O}_S e^{-iHt}\,. \tag{5.23}$$

There is a third, intermediate picture, known as *interaction* or *Dirac picture*, where both the observables and the states evolve in time, the former with the *free* Hamiltonian, and the latter essentially with the interaction part only,

$$\text{Dirac (interaction picture):}\quad |\psi(t)\rangle_I = e^{iH_0t}e^{-iHt}|\psi(0)\rangle_S\,, \qquad \mathcal{O}_I(t) = e^{iH_0t}\mathcal{O}_S e^{-iH_0t}\,. \tag{5.24}$$

Clearly, expectation values (and thus the physics) is independent of the picture one uses,

$${}_S\langle\phi(t)|\mathcal{O}_S|\psi(t)\rangle_S = {}_H\langle\phi|\mathcal{O}_H(t)|\psi\rangle_H = {}_I\langle\phi(t)|\mathcal{O}_I(t)|\psi(t)\rangle_I\,, \tag{5.25}$$

as can be explicitly verified.

5.2 Cross sections

As we discussed already in Section 1.6, the rate at which scattering events happen in a scattering experiment is governed by the total cross section σ of the process: if a beam of particles of flux Φ (particles per unit area per unit time) is directed on a target of N_t particles, then the number of events per unit time is given by the formula $\Delta N_{\text{events}}/\Delta t = \sigma N_t \Phi$. Conversely, to measure the total cross section of a process one counts scattering event for the given experimental setup and derives

$$\sigma = \frac{\Delta N_{\text{events}}}{\Delta t N_t \Phi} \,. \tag{5.26}$$

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

$$\Delta\sigma_\alpha(\xi) = \frac{\Delta N_{\text{events}}(\alpha,\xi)}{\Delta t \Delta\xi N_t \Phi}\Delta\xi \,, \tag{5.27}$$

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,

$$\frac{d\sigma_\alpha(\xi)}{d\xi} = \frac{dN_{\text{events}}(\alpha,\xi)}{dt d\xi N_t \Phi} \,. \tag{5.28}$$

A comment about the notation. If one focuses on a certain particle content in the final state, and chooses to classify the final state according to, e.g., the three momentum components of particle 1, one would write for the corresponding differential cross section something like

$$\frac{d\sigma}{dp_1^{(1)} dp_2^{(1)} dp_3^{(1)}} = \frac{d\sigma}{d^3p^{(1)}} \,. \tag{5.29}$$

The differentials at the denominator should not be understood as derivatives taken on the function σ, but rather as the variables according to which we are distributing our events. If we wanted to know how many events per unit time to expect in corresponding to a certain finite interval of the final-state variables, then we would have to integrate over these variables. In the example above, and asking for the momentum components to satisfy $|p_i^{(1)}| \leq C$, one would have ($\mathcal{D} \equiv [-C,C]\times[-C,C]\times[-C,C]$)

$$\left.\frac{dN_{\text{events}}}{dt}\right|_{|p_i^{(1)}|\leq C} = N_t\Phi \int_{\mathcal{D}} d^3p \frac{d\sigma}{d^3p^{(1)}} \,, \tag{5.30}$$

where N_t and Φ depend on the experiment setup.

The cross section is directly related to the transition probability P_{fi} from an initial state $|\phi_i\rangle$ to a final state 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 do 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 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. Consider now the expression for a differential cross section. The number of events divided by the number of target particles and the number of beam particles that have come in contact is the probability that a scattering event takes place. If we (ideally) use a target with a single scatterer and a beam with a single particle, ΔN_{events} becomes simply the probability P_{fi} of a single scattering event starting with a specific initial state with specific final state properties, and so we can write

$$\Delta\sigma = \frac{P_{fi}}{\Delta t \Phi(N_b = 1)} \,. \tag{5.31}$$

As we have said above, what is measured in experiments is the transition probability averaged over slightly different initial states, so it is not a priori clear what P_{fi} we should use here. 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 experiment 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.

Eq. (5.28) establishes the starting point for relating theory and experiments: in principle we can compute the probabilities P_{fi} from our theories. Quantum mechanics in fact tells us that

$$P_{fi} = \frac{|\langle\phi_f|S|\phi_i\rangle|^2}{\langle\phi_f|\phi_f\rangle\langle\phi_i|\phi_i\rangle} \,. \tag{5.32}$$

In the limit we are interested in, however, that in which initial and final states are momentum eigenstates, we have to deal with non-normalisable states, for which the formula above makes no sense. Another problem is that, since S commutes with energy and momentum, it can certainly be written as

$$S = \mathbf{1} + i(2\pi)^4\delta^{(4)}(P_f - P_i)\mathcal{M} \,, \tag{5.33}$$

where we have made explicit the presence of the no-scattering term $\mathbf{1}$. The formula above, Eq. (5.32), then contains the square of a Dirac delta, which again makes no sense. To circumvent these problems we regularise the expression above by putting our system in a finite four-dimensional box, of time extent T and spatial volume $V = L^3$, and periodic boundary conditions. In this way the allowed momenta become discrete, $p_j = \frac{2\pi k_j}{L}$ with integer k_j, $|k_j| \leq L/2$, and the corresponding eigenstates are normalisable. Moreover, the four-momentum conserving Dirac-delta function in S gets replaced by a Kronecker delta, which can be squared with no complication. However, we have to make sure that when taking the limit of infinite volume everything will be normalised properly. Is is customary for (infinite-volume) momentum eigenstates to employ the relativstic invariant normalisation $\langle\vec{p}\,'|\vec{p}\rangle = (2\pi)^3 2p^0\delta^{(3)}(\vec{p}\,' - \vec{p})$. This fixes also the normalisation of $\mathcal{M}$ in Eq. (5.33). In order for the finite-volume eigenstates $|\vec{p}\rangle_V$ to be

relativistically normalised in the infinite-volume limit we need to impose ${}_V\langle\vec{p}'|\vec{p}\rangle_V = 2p^0 V\delta^{(3)}_{\vec{p}',\vec{p}}$, where the delta is now a Kronecker delta. Indeed, since there is one admissible value of momentum in a cube of volume $V/(2\pi)^3$, the density of modes is $(2\pi)^3/V$ and so $\sum_{\vec{p},V} \to (2\pi)^3/V \int d^3p$ in the infinite-volume limit; from this, the relation between the delta functions follows. Moreover, to get Eq. (5.33) back in the infinite-volume limit, we have to set

$$S_{T,V} = \mathbf{1}_{T,V} + iTV\delta^{(4)}_{P_f,P_i}\mathcal{M}_{T,V}\,. \tag{5.34}$$

Indeed, since $TV\delta^{(4)}_{P_f,P_i} \to (2\pi)^4\delta^{(4)}(P_f - P_i)$ by the same argument used above, so that $\mathcal{M}_{T,V} \to \mathcal{M}$ as $T, V \to \infty$.

We are now ready for the derivation. Let us denote with ΔP_{fi} the probability to obtain a final state of particles with prescribed momenta $\vec{p}_i'$, starting from an initial state of two particles with momenta $\vec{p}_1$ and $\vec{p}_2$. We consider only final states that differ from the initial one, $f \neq i$. Having in mind the infinite-volume limit, we denote with $\Delta^3 p_i' = (2\pi)^3/V$ the size of the "unit cell" in the phase space of final particle i. Plugging the state normalisation and the regularised relation Eq. (5.34) in Eq. (5.32) we find, after dropping the no-scattering term,

$$\Delta P_{fi} = (TV)^2\delta^{(4)}_{P_f,P_i}\frac{|\mathcal{M}_{T,V}|^2}{4p_1^0p_2^0V^2}\prod_j \frac{1}{2p_j'^0V}\frac{\Delta^3 p_i' V}{(2\pi)^3}\,. \tag{5.35}$$

For large V, T we can replace $\mathcal{M}_{T,V}$ with $\mathcal{M}$ and combine one factor VT with the Kronecker delta to obtain a Dirac delta. We get

$$\Delta P_{fi} \longrightarrow \frac{T}{V}\frac{|\mathcal{M}_{T,V}|^2}{4p_1^0p_2^0}(2\pi)^4\delta^{(4)}(P_f - P_i)\prod_j \frac{1}{2p_j'^0}\frac{\Delta^3 p_i'}{(2\pi)^3}\,. \tag{5.36}$$

Since we are looking at an elementary process involving only two particles, there is only one particle in the target and one particle in the beam, which has therefore a flux $\Phi = \frac{v}{V}$, where v is the beam particle velocity. The time it takes for the process to complete is T. The differential cross section corresponding to ΔP_{fi} is thus $\Delta\sigma = \Delta P_{fi}/(T\Phi) = \Delta P_{fi}V/(Tv)$. Using Eq. (5.36), we find

$$\Delta\sigma = \frac{\Delta P_{fi}}{T\Phi} = (2\pi)^4\delta^{(4)}(P_f - P_i)\frac{|\mathcal{M}_{fi}|^2}{4p_1^0p_2^0v}\prod_j \frac{\Delta^3 p_i'}{2p_j'^0(2\pi)^3}\,. \tag{5.37}$$

All factors of V, T cancelled, and we can finally take the limit $T, V \to \infty$ to conclude that

$$d\sigma = (2\pi)^4\delta^{(4)}(P_f - P_i)\frac{|\mathcal{M}_{fi}|^2}{4p_1^0p_2^0v}\prod_j \frac{d^3 p_i'}{2p_j'^0(2\pi)^3} = (2\pi)^4\delta^{(4)}(P_f - P_i)\frac{|\mathcal{M}_{fi}|^2}{4p_1^0p_2^0v}d\Phi^{(n)}\,. \tag{5.38}$$

Here $d\Phi^{(n)}$ is the infinitesimal invariant-volume element of the n-particle phase space,

$$d\Phi^{(n)} = \prod_j \frac{d^3 p_i'}{2p_j'^0(2\pi)^3}(2\pi)^4\delta^{(4)}(P_f - P_i)\,. \tag{5.39}$$

The discussion here was done in the laboratory frame in which one of the particles is at rest, but we would rather have a Lorentz-invariant definition of the cross section. To this end, notice that in the lab frame $p_1^0p_2^0v = E_1vm_2 = |\vec{p}_1|m_2$, which is equal to the following Lorentz-invariant

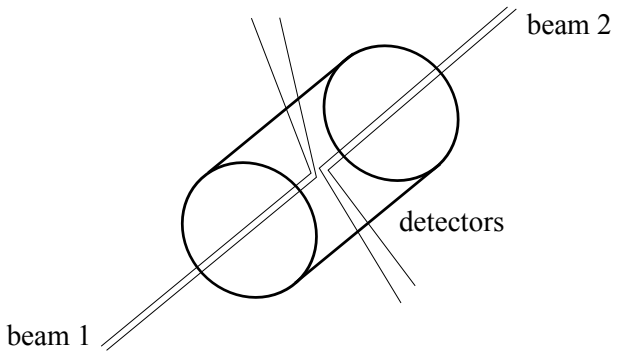

Figure 29: Collider experiment.

quantity evaluated in the lab frame, $|\vec{p}_1|m_2 = \sqrt{(p_1 \cdot p_2)^2 - p_1^2 p_2^2}$. To make Eq. (5.38) manifestly invariant we then write it as

$$\frac{d\sigma}{d\Phi^{(n)}} = \frac{|\mathcal{M}_{fi}|^2}{4\sqrt{(p_1 \cdot p_2)^2 - p_1^2 p_2^2}} \,. \tag{5.40}$$

The invariance is manifest since if we are able to construct a Lorentz-invariant theory with a Lorentz-invariant S-operator, then the matrix element $\mathcal{M}_{fi}$ will be Lorentz-invariant thanks to our choice of a relativistically invariant normalisation for the momentum eigenstates.

For completeness, let us briefly discuss collider experiments. In this case there are two beams of particles 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} 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 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,[37] 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. The number of events per unit time will then be

$$\frac{\Delta N_{\text{events}}}{\Delta t} = \frac{2}{T} \frac{N_{B1} N_{B2} N_{b1} N_{b2}}{A_b} \, \sigma = \mathcal{L}\sigma \,, \tag{5.41}$$

where $\mathcal{L}$ is the *luminosity* delivered by the collider. The analogue of Eq. (5.26) reads then

$$\sigma = \frac{1}{\mathcal{L}} \frac{\Delta N_{\text{events}}}{\Delta t} \,, \tag{5.42}$$

and similarly for the differential cross section.

To make further progress, a detailed theory to compute $\mathcal{M}_{fi}$ is needed. This is the purpose of quantum field theory.

[37] 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.

6 Basics of Quantum Field Theory

It is a fact of Nature that particles can be created or destroyed, not just in the way in which a proton and an electron are created by the dissolution of a hydrogen atom, but also out of the blue as in, say, the pion decay $\pi^+ \to \mu^+ \nu_\mu$. The formalism of quantum mechanics does not allow to describe such processes: to each particle in the game corresponds a wave function, how can that just appear or disappear? This limitation of quantum mechanics is not really a problem in the non-relativistic, low-energy regime where this theory applies. Indeed, to create an extra electron-positron pair starting from a proton-antiproton pair, we need for the proton energy in the CM frame to satisfy $E \geq m_p + m_e$, which results into $\beta \gtrsim 0.1$, i.e., proton and antiproton should travel at $c/10$ to produce a pair of the lightest leptons. At high energies, however, particle production and annihilation become important, and we have to devise a method to describe these processes. Such method must take into account the principles of special relativity, so must lead to Lorentz-invariant results, and should also be quantum-mechanical in Nature, if we are to describe the microscopic world.

6.1 Fock space

The first step is to develop the kinematics, starting from the simplest case of free particles. Free many-particle states are not only an easy starting point for the theoretical development, but they are also of practical relevance since they describe the initial and final states of scattering processes. Let us consider a system of non-interacting bosons, more specifically one type of spinless bosons of mass m. The most general state of such a system is obtained from the linear superposition of states with an arbitrary number n of particles with definite momenta, i.e., eigenstates of the momentum operator. A basis for the Hilbert space of this system is thus $\{|\vec{p}_1, \ldots, \vec{p}_n\rangle\}_{n=0,1,\ldots,\infty}$, where the $n = 0$ state $|0\rangle$ is the *vacuum state*, where no particle is present. It is an experimental fact that particles of the same type are indistinguishable, and in the case of bosons their quantum state vector must be left invariant by any permutation P of the particle labels,

$$|\vec{p}_{\mathrm{P}(1)}, \ldots, \vec{p}_{\mathrm{P}(n)}\rangle = |\vec{p}_1, \ldots, \vec{p}_n\rangle \,, \tag{6.1}$$

i.e., they have to obey *Bose-Einstein statistics*. Formally, these states are obtained from the one-particle states $|\vec{p}\rangle$ by fully symmetrising their n-fold tensor product,

$$|\vec{p}_1, \ldots, \vec{p}_n\rangle = \frac{1}{n!} \sum_{\mathrm{P}} |\vec{p}_{\mathrm{P}(1)}\rangle \otimes \ldots \otimes |\vec{p}_{\mathrm{P}(n)}\rangle \,. \tag{6.2}$$

The space generated by this basis is called the *Fock space* of the system. Since energy and momentum are related by the dispersion relation $E^2 = \vec{p}^{\,2} + m^2$, the one-particle vectors are actually eigenvectors of the four-momentum operator P^μ,

$$P^\mu |\vec{p}\rangle = p^\mu |\vec{p}\rangle \,, \tag{6.3}$$

where $p^0 = \sqrt{\vec{p}^{\,2} + m^2}$. It follows of course that the n-particle states are four-momentum eigenvectors as well.

Basis vectors are usually normalised according to

$$\langle \vec{p}\,' | \vec{p} \rangle = (2\pi)^3 2 p_0 \delta^{(3)}(\vec{p}\,' - \vec{p}) \,, \tag{6.4}$$

for one-particle states, and with the appropriate symmetric generalisation for many-particle states,

$$\langle \vec{p}\,'_1, \ldots, \vec{p}\,'_{n'} | \vec{p}_1, \ldots, \vec{p}_n \rangle = \delta_{n'n} \sum_{\mathrm{P}} \prod_{j=1}^{n} (2\pi)^3 2p_j^0 \delta^{(3)}(\vec{p}\,'_{\mathrm{P}(j)} - \vec{p}_j)\,. \tag{6.5}$$

The vacuum state is instead normalised to 1,

$$\langle 0|0\rangle = 1\,. \tag{6.6}$$

While eigenvectors corresponding to different eigenvalues have automatically a vanishing scalar product, the diagonal elements in Eq. (6.4) are somewhat arbitrary. The choice made in Eq. (6.4) is the *relativistic invariant normalisation*, which takes the same form in any inertial reference frame. Indeed, since Lorentz symmetry will be implemented through unitary transformations on our Hilbert space, and since they will transform the four-momentum operator P^μ as a four-vector, we must have $U(\Lambda)|\vec{p}\rangle = c_{\vec{p}}(\Lambda)|\Lambda\vec{p}\rangle$, with obvious meaning of the notation, and with $c_{\vec{p}}(\Lambda)$ some phase. It then follows, if we set $\langle \vec{p}\,'|\vec{p}\rangle = N(\vec{p})\delta^{(3)}(\vec{p}\,' - \vec{p})$, that

$$\begin{aligned} N(\vec{p})\delta^{(3)}(\vec{p}\,' - \vec{p}) = \langle \vec{p}\,'|\vec{p}\rangle &= \langle \vec{p}\,'|U(\Lambda)^\dagger U(\Lambda)|\vec{p}\rangle = c_{\vec{p}\,'}(\Lambda)^* c_{\vec{p}}(\Lambda)\langle \Lambda\vec{p}\,'|\Lambda\vec{p}\rangle \\ &= N(\Lambda\vec{p})\delta^{(3)}(\Lambda\vec{p}\,' - \Lambda\vec{p})\,. \end{aligned} \tag{6.7}$$

Since $\vec{p}\,' = \vec{p}$ implies $p'^0 = p^0$, $\Lambda\vec{p}\,' = \Lambda\vec{p}$ implies the four-vector relation $\Lambda p' = \Lambda p$, which is true if and only if $p' = p$. The last delta function in Eq. (6.7) is thus proportional to $\delta^{(3)}(\vec{p}\,' - \vec{p})$. To find the proportionality constant, recall that a generic (proper orthocronous) Lorentz transformation is obtained combining rotations and boosts along, say, the 1 direction. For rotations, one autoamtically has invariance of the delta function, so we can focus on the 1-boosts only. We have $(\Lambda p)^0 = \gamma(p^0 - \beta p^1)$, $(\Lambda p)^1 = \gamma(p^1 - \beta p^0)$, and so

$$\begin{aligned} \delta((\Lambda p)'^1 - (\Lambda p)^1) = \delta(\gamma[(p'^1 - p^1) - \beta(p'^0 - p^0)]) &= \left|\frac{\partial\gamma(p^1 - \beta p^0)}{\partial p^1}\right|^{-1} \delta(p'^1 - p^1) \\ = p^0|\gamma(p^0 - \beta p^1)|^{-1}\delta(p'^1 - p^1) &= \frac{p^0}{(\Lambda p)^0}\delta(p'^1 - p^1)\,. \end{aligned} \tag{6.8}$$

From this we conclude that in general we must have

$$\frac{N(\Lambda\vec{p})}{(\Lambda p)^0} = \frac{N(\vec{p})}{p^0}\,. \tag{6.9}$$

The simplest way to achieve this is to set $N(\vec{p}) \propto p^0$. The same argument shows the invariance under Lorentz transformations of the integration measure

$$d\Omega_p \equiv \frac{d^3p}{(2\pi)^3 2p^0}\,. \tag{6.10}$$

We now define the *creation* and *annihilation operators* as follows. The creation operator $a(\vec{p})^\dagger$ is defined via

$$a(\vec{p})^\dagger |\vec{p}_1, \ldots \vec{p}_n\rangle \equiv |\vec{p}, \vec{p}_1, \ldots \vec{p}_n\rangle\,. \tag{6.11}$$

the annihilation operator $a(\vec{p})$ is defined as the adjoint of $a(\vec{p})^\dagger$. It is easy to show, using the definition Eq. (6.11) and the relations Eq. (6.5) that

$$a(\vec{p})|\vec{p}_1,\dots\vec{p}_n\rangle = \sum_{j=1}^{n}(2\pi)^3 2p_j^0\delta^{(3)}(\vec{p}_j-\vec{p})|\vec{p}_1,\dots,\vec{p}_{j-1},\vec{p}_{j+1},\dots\vec{p}_n\rangle\,. \tag{6.12}$$

These operators allow us to change as we please the particle content of a state. In particular, since we cannot remove particles from the vacuum, we must have $a(\vec{p})|0\rangle = 0\ \forall\vec{p}$. This is the unique state in Fock space for which this happens, so this feature singles it out. Any other state can be obtained from the vacuum by repeated application of creation operators,

$$|\vec{p}_1,\dots\vec{p}_n\rangle \equiv a(\vec{p}_1)^\dagger\dots a(\vec{p}_n)^\dagger|0\rangle\,. \tag{6.13}$$

This allows us to give an equivalent characterisation of Fock space. First of all, notice that the symmetry under permutations of the indices of the multiparticle states implies that

$$[a(\vec{p}),a(\vec{q})] = [a(\vec{p})^\dagger,a(\vec{q})^\dagger] = 0\,. \tag{6.14}$$

Next, using Eqs. (6.11) and (6.12) and the relations Eq. (6.5), one finds that

$$[a(\vec{p}),a(\vec{q})^\dagger] = (2\pi)^3 2p^0\delta^{(3)}(\vec{p}'-\vec{p})\,. \tag{6.15}$$

Fock space is the basis for an irreducible representation of the commutator algebra Eq. (6.14)–(6.15) in which the number operator,

$$N = \int\frac{d^3p}{(2\pi)^3 2p^0}\,a(\vec{p})^\dagger a(\vec{p})\,. \tag{6.16}$$

is diagonal. Indeed, N is a positive Hermitian operator, which we can diagonalise; since application of the annihilation operator on an eigenvector leads to a new eigenvector with eigenvalue decreased by one, as a consequence of Eq. (6.15), at some point we must get zero. The state which is annihilated by all annihilation operators is unique if the representation is irreducible. The construciton of the other states is then straightforward.

The creation and annihilation operators allow us to write a compact expression for the energy and momentum operators. As we have mentioned above, the eigenvalue of the number operator decreases by one under application of an annihilation operator, and correspondingly it increases by one when we apply a creation operator. This follows from the following results,

$$\begin{aligned}
[a(\vec{p})^\dagger a(\vec{p}),a(\vec{q})] &= [a(\vec{p})^\dagger,a(\vec{q})]a(\vec{p}) = -(2\pi)^3 2p^0\delta^{(3)}(\vec{q}-\vec{p})a(\vec{p})\,,\\
[a(\vec{p})^\dagger a(\vec{p}),a(\vec{q})^\dagger] &= a(\vec{p})^\dagger[a(\vec{p}),a(\vec{q})^\dagger] = (2\pi)^3 2p^0\delta^{(3)}(\vec{q}-\vec{p})a(\vec{p})^\dagger\,,
\end{aligned} \tag{6.17}$$

which characterises $a(\vec{p})^\dagger a(\vec{p})$ as a number-density operator. Given that energy and momentum are additive quantities for free particles, we immediately find that the total four-momentum operator is given by

$$P^\mu = \int\frac{d^3p}{(2\pi)^3 2p^0}\,p^\mu a(\vec{p})^\dagger a(\vec{p})\,. \tag{6.18}$$

The numerator in the integration measure in Eqs. (6.16) and Eq. (6.18) is fixed by our choice of normalisation. The integration measure,

$$d\Omega_p \equiv \frac{d^3p}{(2\pi)^3 2p^0}\,, \tag{6.19}$$

is the one-particle invariant phase-space measure. In general, for any additive quantum number $f(\vec{p})$ we have that the corresponding operator F reads

$$F = \int \frac{d^3p}{(2\pi)^3 2p^0}\, f(\vec{p}) a(\vec{p})^\dagger a(\vec{p})\,. \tag{6.20}$$

There is not much left to say about the Fock space formalism. It is easy to extend it to the case of several different types of boson, by introducing the corresponding creation and annihilation operators and imposing the operators for particles of different types commute with each other. The extension of fermions requires instead that the Fermi-Dirac statistics be imposed on the states, which requires the introduction of a minus sign in the state vector when we exchange the labels of any two particles. This reflects on the creation and annihilation operators, which must now satisfy the following *anticommutation relations*,

$$\begin{aligned} \{a(\vec{p}), a(\vec{q})\} &= \{a(\vec{p})^\dagger, a(\vec{q})^\dagger\} = 0\,, \\ \{a(\vec{p}), a(\vec{q})^\dagger\} &= (2\pi)^3 2p^0 \delta^{(3)}(\vec{p}' - \vec{p})\,. \end{aligned} \tag{6.21}$$

We note in passing that these relation lead to Pauli exclusion principle, as no more than one particle per state can be created. Operators corresponding to different types of fermions are taken to anticommute with each other. Finally, bosonic and fermionic operators commute with each other.

6.2 The second quantisation

Quantum mechanics is formulated in a non-relativistic setting, and so has to be properly modified to be compatible with the tenets of special relativity. In particular, the Schrödinger equation governing the time evolution of the wave function is not relativistically covariant. It is not difficult to write down a relativistic analogue of that equation (and that is actually what Schrödinger did first), and we did that already in Section 1.8: this is the Klein-Gordon equation,

$$(\Box + m^2)\phi(x) = 0\,. \tag{6.22}$$

This equation, however, leads to inconsistencies (and that is why Schrödinger dropped it initially, focussing on its non-relativistic limit). First of all, being second-order in time, as well as in space, it admits negative-energy solutions, which would lead to instabilities, as the system could always transition to a lower energy state. Secondly, the only probability current with the right symmetry properties under Lorentz transformations leads to a probability density which is not positive-definite.

There is a deep reason for the failure of relativistic quantum mechanics. Quantum mechanics (relativistic or not) describes physical systems in terms of wave functions, which implicitly assumes that the number and type of particles remains constant in time. This does not reflect the actual state of things: particles can be created and destroyed in Nature, essentially because of the mass/energy equivalence relation of special relativity. The appropriate generalisation of quantum mechanics to the relativistic regime must take this into account.

In order to proceed, we have to rethink the role of the solution to relativistic equations like Eq. (6.22). The wave function of quantum mechanics is technically a complex field, i.e., a complex-valued function defined at every point in space. It is essentially a probability field, as its modulus square gives the probability density function for the corresponding particle to

be found at a given point in space at a given time. By contrast, the classical electromagnetic field describes the strength and orientation of the electric and magnetic field at a given point in space. However, the quantum hypothesis of Planck (1900) and Einstein (1905) amounts to assuming that the energy stored in each normal mode $(\omega, \vec{k})$ of the field cannot be arbitrary, but only an integer multiple of the energy quantum $\hbar\omega$. These quanta are nothing but the photons, which have all the features of (massless) particles. The quantised version of the electromagnetic field should therefore describe the excitation or de-excitation of the normal modes, or equivalently the creation and annihilation of photons. By analogy, we can then think of "quantising the wave function", i.e., the probability field, of a particle, so that it will describe the creation and annihilation of the corresponding type of particles. This procedure is called *second quantisation.*[38]

Having made the case for it, let us now discuss how second quantisation is done. We begin by associating a field to a particle type, rather than a wave function to a given particle. As we are still classical at this stage, this is just renaming the object $\phi = \phi(x)$ appearing in the Klein-Gordon equation Eq. (6.22). We now solve the equation going over to momentum space, setting

$$\phi(x) = \int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x}\tilde{\phi}(p)\,. \tag{6.23}$$

In momentum space we find

$$\begin{aligned}(p^2 - m^2)\tilde{\phi}(p) &= 0 \Rightarrow \\ \tilde{\phi}(p) &= 2\pi\delta(p^2 - m^2) f(p^0, \vec{p}) \\ &= \frac{2\pi}{2\varepsilon(\vec{p})}\left\{\delta(p^0 - \varepsilon(\vec{p}))f(\varepsilon(\vec{p}), \vec{p}) + \delta(p^0 + \varepsilon(\vec{p}))f(-\varepsilon(\vec{p}), \vec{p})\right\}\,,\end{aligned} \tag{6.24}$$

where $\varepsilon(\vec{p}) \equiv \sqrt{\vec{p}^2 + m^2}$. Performing the integration,

$$\begin{aligned}\phi(x) &= \int \frac{d^3p}{(2\pi)^3 2\varepsilon(\vec{p})}\left\{e^{-i(\varepsilon(\vec{p})x^0 - \vec{p}\cdot\vec{x})} f(\varepsilon(\vec{p}), \vec{p}) + e^{i(\varepsilon(\vec{p})x^0 + \vec{p}\cdot\vec{x})} f(-\varepsilon(\vec{p}), \vec{p})\right\} \\ &= \int \frac{d^3p}{(2\pi)^3 2\varepsilon(\vec{p})}\left\{e^{-i(\varepsilon(\vec{p})x^0 - \vec{p}\cdot\vec{x})} f(\varepsilon(\vec{p}), \vec{p}) + e^{i(\varepsilon(\vec{p})x^0 - \vec{p}\cdot\vec{x})} f(-\varepsilon(\vec{p}), -\vec{p})\right\}\,.\end{aligned} \tag{6.25}$$

We now set $d\Omega_p = d^3p/[(2\pi)^3 2p^0]$, where from now on we understand that $p^0 = \varepsilon(\vec{p})$, and furthermore we set $a(\vec{p}) = f(\varepsilon(\vec{p}), \vec{p})$ and $b(\vec{p})^\dagger = f(-\varepsilon(\vec{p}), -\vec{p})$, and write

$$\phi(x) = \int d\Omega_p \left\{a(\vec{p})e^{-ip\cdot x} + b(\vec{p})^\dagger e^{ip\cdot x}\right\}\,. \tag{6.26}$$

So far we have just introduced some fancy notation for the classical solution of the Klein-Gordon equation. The quantisation step comes now, when promoting the amplitude *functions* $a(\vec{p})$ and $b(\vec{p})^\dagger$, corresponding to the various normal modes, to *operators* that control the number of excitations corresponding to that normal mode, i.e., the number of particles with that momentum. We have already met the appropriate operators for this task: they are the creation and

[38] I mention it here mostly for historical and pedagogical reasons. A better understanding of quantum fields is achieved ignoring completely the idea of "quantising the wave function".

annihilation operators of the previous subsection. We then set

$$\begin{aligned}
[a(\vec{p}), a(\vec{q})^\dagger] = [b(\vec{p}), b(\vec{q})^\dagger] = (2\pi)^3 2p^0 \delta^{(3)}(\vec{p}' - \vec{p})\,, \\
[a(\vec{p}), a(\vec{q})] = [a(\vec{p})^\dagger, a(\vec{q})^\dagger] = [b(\vec{p}), b(\vec{q})] = [b(\vec{p})^\dagger, b(\vec{q})^\dagger] = 0\,, \\
[a(\vec{p}), b(\vec{q})] = [a(\vec{p}), b(\vec{q})^\dagger] = [a(\vec{p})^\dagger, b(\vec{q})] = [a(\vec{p})^\dagger, b(\vec{q})^\dagger] = 0\,.
\end{aligned} \tag{6.27}$$

These operators act on the Fock space of the particles, which is built by their repeated application on the vacuum state $|0\rangle$, defined by $a(\vec{p})|0\rangle = b(\vec{p})|0\rangle = 0$, $\forall \vec{p}$.

Let us stop for a second and assess the situation. By means of a heuristic argument we have made the wave function of a free system into a quantum field operator built out of creation and annihilation operators. By itself this is a completely useless gimmick: we already knew how to describe free particles, with no need for the field operator. It is possible to show that at least we are not losing anything, i.e., the description in terms of $\phi(x)$ is equivalent to that in terms of creation and annihilation operators. This follows from the identities

$$\begin{aligned}
a(\vec{p}) &= \int d^3x\, e^{ip\cdot x} i\overset{\leftrightarrow}{\partial}_0 \phi(x)\,, & b(\vec{p}) &= \int d^3x\, e^{ip\cdot x} i\overset{\leftrightarrow}{\partial}_0 \phi(x)^\dagger\,, \\
a(\vec{p})^\dagger &= -\int d^3x\, e^{ip\cdot x} i\overset{\leftrightarrow}{\partial}_0 \phi(x)^\dagger\,, & b(\vec{p})^\dagger &= -\int d^3x\, e^{ip\cdot x} i\overset{\leftrightarrow}{\partial}_0 \phi(x)\,,
\end{aligned} \tag{6.28}$$

where $f\overset{\leftrightarrow}{\partial}_0 g \equiv f(\partial_0 g) - (\partial_0 f)g$. Also, we started trying to obtain a quantum relativistic description of one type of particle, and we ended up with two sets of creation and annihilation operators. In fact, there is a priori no reason to assume any relation between the a-type and b-type operators: the classical field ϕ is generally a complex field, so the positive-frequency component $f(\varepsilon(\vec{p}), \vec{p})$ and the negative-frequency component $f(\varepsilon(-\vec{p}), -\vec{p})$ are independent. If we assumed that the classical field were real, then the two components would be related by complex conjugation, and after quantisation this would lead to identify a-type and b-type operators, making the quantum field Hermitian. This is, however, a generally unmotivated restriction. therefore, in general we will be describe two types of particles with the same mass m.

To understand the importance of the quantum field, and the meaning of the two sets of creation/annihilation operators, we have to think ahead, and imagine what we would have to do when trying to build an interacting theory in order to describe the real world. Any attempt at building such a theory must take into account the constraints coming from experience, which include *Lorentz* and *translation invariance*, and *locality* and *causality* of the interaction. Here we start to see the utility of the quantum field. In principle, we could try to built interactions out of creation and annihilation operators, but it would be very difficult to ensure locality of the interactions while working in momentum space. The field $\phi(x)$ living in coordinate space is much better suited for this. Concerning the symmetry properties of the field, these are inherited from those of the particle states. Indeed, Wigner's theorem ensures that it will be possible to realise Lorentz transformations and translations as unitary operators on the Hilbert space of the particles. For the momentum eigenstates of scalar particles, the effect of these unitary operators is easy to determine: looking at the same state in different reference frames we would assign t oit two different values of momenutm related by the Lorentz transformation connecting the two frames. In formulae,

$$U(\Lambda)|\vec{p}\rangle = |\Lambda\vec{p}\rangle\,, \tag{6.29}$$

up to phases that can always be set to 1 by properly redefining the states. The effect of translations is even easier to determine: as these are generated by the four-momentum operator,

one has

$$U(a)|\vec{p}\rangle = e^{-ia\cdot P}|\vec{p}\rangle = e^{-ia\cdot p}|\vec{p}\rangle\,. \tag{6.30}$$

Multiparticle states transform just like the direct product of one-particle states. From this follows that the vacuum state is invariant under both types of transformations,[39]

$$U(\Lambda)|0\rangle = U(a)|0\rangle = |0\rangle\,. \tag{6.31}$$

From Eqs. (6.29)–(6.31) one easily determines the transformation laws of the creation operators:

$$U(\Lambda)^\dagger a(\vec{p})^\dagger U(\Lambda) = a(\Lambda^{-1}\vec{p})^\dagger\,, \qquad U(a)^\dagger a(\vec{p})^\dagger U(a) = e^{ia\cdot p}a(\vec{p})^\dagger\,, \tag{6.32}$$

and similarly for $b(\vec{p})^\dagger$. Those of the annihilation operators follow by taking the Hermitian conjugate if Eq. (6.32). From these relations, and taking into account that the integration measure is invariant under Lorentz transformations, one finds that

$$\begin{aligned} U(\Lambda)^\dagger\phi(x)U(\Lambda) &= \phi(\Lambda^{-1}x)\,,\\ U(a)^\dagger\phi(x)U(a) &= \phi(x+a)\,. \end{aligned} \tag{6.33}$$

This is precisely how we expect a scalar field to transform under a change of reference frame: if we define a new set of coordinates via $x' = \Lambda x$, the field in the new frame at the point with new coordinates x' must to be equal to the field in the old frame at the same point, which in that frame has coordinates x. In formulae, $\phi(x) = \phi'(x') = U(\Lambda)^\dagger\phi(x')U(\Lambda) = U(\Lambda)^\dagger\phi(\Lambda x)U(\Lambda)$, which is equivalent to the first equation in Eq. (6.33). The same argument applies to translations: changing coordinates to $x' = x - a$ by translating the origin of the reference frame by a, we must have $\phi(x) = \phi'(x') = U(a)^\dagger\phi(x')U(a) = U(a)^\dagger\phi(x-a)U(a)$, which is equivalent to the second equation in Eq. (6.33). The simplicity of the transformation laws Eq. (6.33) means that it will be easy to construct operators with prescribed transformation properties out of $\phi(x)$, so that it will not take much of an effort to build a Lorentz-invariant theory.

An extremely useful property of the field $\phi(x)$ is that it obeys the so-called *microcausality relations*, i.e., the operators provided by the field, or its Hermitian conjugate, or any of its derivatives, at points x and y will commute if x and y are spacelike-separated. The importance of this property lies in the fact that if we built observables out of ϕ, $\phi^\dagger$, and their derivatives, they will automatically commute for spacelike separations, reflecting the fact that measurements made on the system at spacelike-separated points cannot affect each other (due to the finiteness of the speed of light), since those points are causally disconnected. Moreover, this property will be used to prove the Lorentz-invariance of the S-matrix. Let us now show it. We find from Eq. (6.26) and the commutation relations Eq. (6.27) that $[\phi(x), \phi(y)] = 0$ for any pair of points. Furthermore,

$$\begin{aligned} [\phi(x), \phi(y)^\dagger] &= \int d\Omega_p \int d\Omega_q \left[e^{-i(p\cdot x - q\cdot y)} - e^{i(p\cdot x - q\cdot y)}\right](2\pi)^3 2q^0\delta^{(3)}(\vec{p}-\vec{q})\\ &= \int d\Omega_p \left[e^{-ip\cdot(x-y)} - e^{ip\cdot(x-y)}\right] = \Delta(x-y) - \Delta(y-x)\,. \end{aligned} \tag{6.34}$$

[39] From $\langle\vec{p}|U(\Lambda,a)|0\rangle = e^{ia\cdot p}\langle\Lambda^{-1}\vec{p}|0\rangle = 0$, where $U(\Lambda,a) = U(a)U(\Lambda)$, and its generalisation to multiparticle states, follows that $U(\Lambda,a)|0\rangle = e^{i\varphi_{(\Lambda,a)}}|0\rangle$. The phase factors must realise a one-dimensional representation of the Poincaré group (Lorentz transformations and translations), but the only such representation is the trivial one.

The function $\Delta(x)$,

$$\Delta(x) \equiv \int d\Omega_p \, e^{-ip\cdot x} \,, \tag{6.35}$$

is a Lorentz-invariant function due to the invariance of the integration measure. Furthermore, since a spacelike $x = (x^0, \vec{x})$ can always be transformed to $(0, \vec{x}\,')$ by means of a Lorentz transformation, we have

$$\Delta(x) = \int d\Omega_p \, e^{i\vec{p}\cdot\vec{x}\,'} = \int d\Omega_p \, e^{-i\vec{p}\cdot\vec{x}\,'} = \Delta(-x) \,, \quad \text{if} \;\; x^2 < 0 \,. \tag{6.36}$$

But then from Eq. (6.37) we get

$$[\phi(x), \phi(y)^\dagger] = 0 \,, \quad \text{if} \;\; (x-y)^2 < 0 \,. \tag{6.37}$$

If we take the derivative of Eq. (6.34) with respect to x we find

$$[\partial_\mu \phi(x), \phi(y)^\dagger] = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} [\phi(x + \epsilon\hat{\mu}) - \phi(x), \phi(y)^\dagger] \,. \tag{6.38}$$

If $(x-y)^2 < 0$, then $(x + \epsilon\hat{\mu} - y)^2 = (x-y)^2 + 2\epsilon(x-y)_\mu + \epsilon^2\hat{\mu}^2$ will also be negative for sufficiently small $\epsilon < \epsilon_0$. Both x and $x + \epsilon\hat{\mu}$ will then be spacelike-separated from y, and the commutator will vanish. We then have in general

$$[\partial^x_{\mu_1} \ldots \partial^x_{\mu_m} \phi(x), \partial^y_{\nu_1} \ldots \partial^y_{\nu_n} \phi(y)^\dagger] = 0 \,, \quad \text{if} \;\; (x-y)^2 < 0 \,, \tag{6.39}$$

which establishes the microcausality of the field ϕ.

6.3 Antiparticles

It is now time to explain why in general we should not impose the Hermiticity condition on the field, thus obtaining a single type of particles. Suppose that we want to describe particles which have some nonzero additive charge q which is conserved, i.e., the particles are eigenstates of some Hermitian operator Q that commutes with the Hamiltonian. This includes, e.g., electric charge, baryon number, and strangeness. Let $a(\vec{p})^\dagger$ and $a(\vec{p})$ be the creation and annihilation operators associated to these particles. If $Q|\vec{p}\rangle = q|\vec{p}\rangle$, one can easily determine the commutation relations

$$[Q, a(\vec{p})] = -qa(\vec{p}) \,, \qquad [Q, a(\vec{p})^\dagger] = qa(\vec{p})^\dagger \,. \tag{6.40}$$

If we have only this type of particles, we are forced to set $a(\vec{p}) = ba(\vec{p})$ in the construction of our field, and so

$$[Q, \phi(x)] = -q \int d\Omega_p \, \left\{ a(\vec{p}) e^{-ip\cdot x} - a(\vec{p})^\dagger e^{ip\cdot x} \right\} = -q\tilde{\phi}(x) \,. \tag{6.41}$$

The only way that a Hamiltonian H built out of fields can commute with Q is if it involves the field $\tilde{\phi}(x)$ besides $\phi(x)$. But

$$\begin{aligned} [\phi(x), \tilde{\phi}(y)^\dagger] &= \int d\Omega_p \int d\Omega_q \, [a(\vec{p}) e^{-ip\cdot x} + a(\vec{p})^\dagger e^{ip\cdot x}, a(\vec{p})^\dagger e^{iq\cdot y} - a(\vec{p}) e^{-iq\cdot y}] \\ &= \Delta(x-y) + \Delta(y-x) \,, \end{aligned} \tag{6.42}$$

which does *not* vanish for $(x-y)^2 < 0$. This means that in general we will not be able to build a Hamiltonian that respects microcausality, unless $q = 0$. Hermitian fields are therefore only adequate to describe fully neutral particles. If we want to describe particles that carry any type of charge we have to try $a \neq b$. Assigning q_a to a-type particles and q_b to b-type particles, Eq. (6.41) becomes in this case

$$[Q, \phi(x)] = -\int d\Omega_p \left\{ q_a a(\vec{p}) e^{-ip\cdot x} - q_b b(\vec{p})^\dagger e^{ip\cdot x} \right\} = -\frac{q_a - q_b}{2}\phi(x) - \frac{q_a + q_b}{2}\tilde{\phi}(x)\,, \tag{6.43}$$

and to avoid problems with microcausality we must set $q_b = -q_a$. Particles of type b have all their conserved charges equal in magnitude and opposite in sign to those of particle of type a, and are usually referred to as their *antiparticles*. The existence of antiparticles is therefore forced on us as soon as we subscribe to quantum field theory and we need to describe particles with nonzero charge.

6.4 Hamiltonian formalism

The properties above are nice and all, but as a matter of fact we are still not even sure that we are describing a system of particles. To this end, let us begin by noticing that since spacelike-separated points can be always Lorentz-transformed to equal-time points, it is enough to study the commutators at equal time to see whether they satisfy microcausality or not. We obviously have

$$[\phi(x), \phi(y)]_{\rm ET} = 0\,, \qquad [\phi(x), \partial_0\phi(y)]_{\rm ET} = 0\,, \qquad [\partial_0\phi(x), \partial_0\phi(y)]_{\rm ET} = 0\,, \tag{6.44}$$

and similarly for the Hermitian conjugate of the field,

$$[\phi(x)^\dagger, \phi(y)^\dagger]_{\rm ET} = 0\,, \qquad [\phi(x)^\dagger, \partial_0\phi(y)^\dagger]_{\rm ET} = 0\,, \qquad [\partial_0\phi(x)^\dagger, \partial_0\phi(y)^\dagger]_{\rm ET} = 0\,. \tag{6.45}$$

From Eq. (6.34) we find

$$[\phi(x), \phi(y)^\dagger]_{\rm ET} = \int d\Omega_p \left[e^{i\vec{p}\cdot(\vec{x}-\vec{y})} - e^{-i\vec{p}\cdot(\vec{x}-\vec{y})} \right] = 0\,, \tag{6.46}$$

for all $\vec{x}$, $\vec{y}$. Here the subscript ET stands for "equal time", i.e., $x^0 = y^0$. Taking the derivative with respect to y^0 and then setting $x^0 = y^0$ we find

$$\begin{aligned} [\phi(x), \partial_0\phi(y)^\dagger]_{\rm ET} &= \int d\Omega_p\, ip^0 \left[e^{i\vec{p}\cdot(\vec{x}-\vec{y})} + e^{-i\vec{p}\cdot(\vec{x}-\vec{y})} \right] = i\int d\Omega_p 2p^0 e^{i\vec{p}\cdot(\vec{x}-\vec{y})} \\ &= i\int \frac{d^3p}{(2\pi)^3} e^{i\vec{p}\cdot(\vec{x}-\vec{y})} = i\delta^{(3)}(\vec{x}-\vec{y})\,. \end{aligned} \tag{6.47}$$

This does not contradict the argument above, as the commutator is only nonzero for $\vec{x} = \vec{y}$ at equal times, which is lightlike rather than spacelike separation. Taking the Hermitian conjugate we obtain

$$[\phi(x)^\dagger, \partial_0\phi(y)]_{\rm ET} = i\delta^{(3)}(\vec{x}-\vec{y})\,. \tag{6.48}$$

If we also take the derivative with respect to x^0 we get

$$\begin{aligned} [\partial_0\phi(x), \partial_0\phi(y)^\dagger]_{\rm ET} &= \int d\Omega_p\, (p^0)^2 \left[e^{i\vec{p}\cdot(\vec{x}-\vec{y})} - e^{-i\vec{p}\cdot(\vec{x}-\vec{y})} \right] \\ &= \int d\Omega_p\, (\vec{p}^{\,2} + m^2) \left[e^{i\vec{p}\cdot(\vec{x}-\vec{y})} - e^{-i\vec{p}\cdot(\vec{x}-\vec{y})} \right] = 0\,, \end{aligned} \tag{6.49}$$

due to symmetry. All other commutators will automatically vanish for spacelike separations, because they are obtained from the relations above through spatial derivatives, which keep the separation spacelike. The case of two or more temporal derivatives is also reduced to these via the equations of motion.

All in all there are ten commutation relations, and one can prove that assuming the commutation relations for the fields, one can derive the commutation relations for the creation and annihilation operators, Eq. (6.27). This provides an interesting perspective on the approach we are following. In fact, Eqs. (6.44)–(6.49) are precisely the commutation relations expected for a set of canonically-conjugate variables. Taking the field and its Hermitian conjugate at time $x^0 = 0$ and all space points, $\phi(0,\vec{x})$ and $\phi(0,\vec{x})^\dagger$ as the canonical coordinates and the temporal derivatives at $x^0 = 0$, $\pi(0,\vec{x}) = \partial_0\phi(0,\vec{x})^\dagger$ and $\pi(0,\vec{x})^\dagger = \partial_0\phi(0,\vec{x})$, as the canonical momenta, then Eqs. (6.44)–(6.47) are just the canonical commutation relations, with all coordiantes commuting with each other, all momenta commuting with each other, and the nontrivial commutation relations

$$[\phi(0,\vec{x}),\pi(0,\vec{y})] = i\delta^{(3)}(\vec{x}-\vec{y})\,, \qquad [\phi(0,\vec{x})^\dagger,\pi(0,\vec{y})^\dagger] = i\delta^{(3)}(\vec{x}-\vec{y})\,. \tag{6.50}$$

The equations of motion read

$$\begin{aligned} \partial_0\phi(t,\vec{x}) &= \pi(t,\vec{x})^\dagger\,, \qquad & \partial_0\pi(t,\vec{x}) &= (\vec{\nabla}^2 - m^2)\phi(t,\vec{x})^\dagger\,, \\ \partial_0\phi(t,\vec{x})^\dagger &= \pi(t,\vec{x})\,, \qquad & \partial_0\pi(t,\vec{x})^\dagger &= (\vec{\nabla}^2 - m^2)\phi(t,\vec{x})\,, \end{aligned} \tag{6.51}$$

and we can easily show that they form a consistent Hamiltonian system, i.e., they can be derived from a Hamiltonian by means of Hamilton equations,

$$\dot{q} = \frac{\partial H}{\partial p}\,, \qquad \dot{p} = -\frac{\partial H}{\partial q}\,. \tag{6.52}$$

Up to a constant, the appropriate Hamiltonian is easily found to be

$$H = \int d^3x \left\{\pi(x)^\dagger\pi(x) + [\vec{\nabla}\phi(x)]^\dagger\cdot[\vec{\nabla}\phi(x)] + m^2\phi(x)^\dagger\phi(x)\right\} + \text{const.}\,. \tag{6.53}$$

In order to check whether we are doing the right thing or not, we have to make sure that this Hamiltonian is the Hamiltonian of a system of free spinless bosons. If this is the case, then we have a nice description of this system in terms of (quantum) Hamiltonian mechanics, as specified by the canonical commutation relations and by the Hamiltonian Eq. (6.53). We then plug Eq. (6.26) into Eq. (6.53) to find

$$\begin{aligned} &\int d^3x \left\{\pi(x)^\dagger\pi(x) + [\vec{\nabla}\phi(x)]^\dagger\cdot[\vec{\nabla}\phi(x)] + m^2\phi(x)^\dagger\phi(x)\right\} \\ &= \int d^3x \left\{\partial_0\phi(x)^\dagger\partial_0\phi(x) + [\vec{\nabla}\phi(x)]^\dagger\cdot[\vec{\nabla}\phi(x)] + m^2\phi(x)^\dagger\phi(x)\right\} \\ &= \int d^3x \int d\Omega_p \int d\Omega_q \left\{m^2\left[a(\vec{p})^\dagger e^{ip\cdot x} + b(\vec{p})e^{-ip\cdot x}\right]\left[a(\vec{q})e^{-iq\cdot x} + b(\vec{p})^\dagger e^{iq\cdot x}\right]\right. \\ &\qquad \left. + \sum_\mu (ip_\mu)(-iq_\mu)\left[a(\vec{p})^\dagger e^{ip\cdot x} - b(\vec{p})e^{-ip\cdot x}\right]\left[a(\vec{q})e^{-iq\cdot x} - b(\vec{p})^\dagger e^{iq\cdot x}\right]\right\}\,. \end{aligned} \tag{6.54}$$

The integration over d^3x yields

$$\int d^3x\, e^{i(\pm)_1(p(\mp)_2 q)\cdot x} = (2\pi)^3\delta^{(3)}(\vec{p}(\mp)_2\vec{q})e^{i(\pm)_1(p^0(\mp)_2 q^0)x^0}\,, \tag{6.55}$$

which leads to

$$\begin{aligned}
&\int d^3x \left\{\pi(x)^\dagger\pi(x) + [\vec{\nabla}\phi(x)]^\dagger\cdot[\vec{\nabla}\phi(x)] + m^2\phi(x)^\dagger\phi(x)\right\} \\
&= \int d\Omega_p \,\frac{1}{2p^0}\left\{(m^2+p_0^2+\vec{p}^2)\left[a(\vec{p})^\dagger a(\vec{p}) + b(\vec{p})b(\vec{p})^\dagger\right]\right. \\
&\qquad\qquad \left. +(m^2-p_0^2+\vec{p}^2)\left[a(\vec{p})^\dagger b(\vec{p})^\dagger e^{i2p^0x^0} + a(\vec{p})b(\vec{p})e^{-i2p^0x^0}\right]\right\} \\
&= \int d\Omega_p\, p^0\left[a(\vec{p})^\dagger a(\vec{p}) + b(\vec{p})b(\vec{p})^\dagger\right]\,.
\end{aligned} \tag{6.56}$$

This is almost what we want, except that the b-type operators are in the wrong order. Putting them in the right one would yield

$$b(\vec{p})b(\vec{p})^\dagger = b(\vec{p})^\dagger b(\vec{p}) + [b(\vec{p}), b(\vec{p})^\dagger] = b(\vec{p})^\dagger b(\vec{p}) + (2\pi)^3 2p^0\delta^{(3)}(0)\,, \tag{6.57}$$

which is clearly nonsense. However, ignoring rigour for a second, the second term in Eq. (6.57) is a constant, independent of $\vec{p}$, although an infinite one, and we are free to redefine the Hamiltonian by dropping it. More rigorously, the product of fields at the same spacetime point was singular to begin with. If we compute the vacuum expectation value of the product of ϕ and $\phi^\dagger$ we find

$$\begin{aligned}
\langle 0|\phi(x)\phi(y)^\dagger|0\rangle &= \int d\Omega_p \int d\Omega_q \langle 0|\left[a(\vec{p})e^{-ip\cdot x} + b(\vec{p})^\dagger e^{ip\cdot x}\right]\left[a(\vec{q})^\dagger e^{iq\cdot y} + b(\vec{p})e^{-iq\cdot y}\right]|0\rangle \\
&= \int d\Omega_p \int d\Omega_q\, e^{-ip\cdot x}e^{iq\cdot y}\langle 0|a(\vec{p})a(\vec{q})^\dagger|0\rangle \\
&= \int d\Omega_p \int d\Omega_q\, e^{-ip\cdot x}e^{iq\cdot y}\langle 0|[a(\vec{p}), a(\vec{q})^\dagger]|0\rangle = \int d\Omega_p\, e^{-ip\cdot(x-y)} \\
&= \Delta(x-y)\,,
\end{aligned} \tag{6.58}$$

which is clearly singular as $x \to y$. Having a singular matrix element, the operator product $\phi(x)\phi(y)^\dagger$ is singular as $x \to y$, and so, strictly speaking, the integral appearing in Eq. (6.53) is not well defined, and so is the Hamiltonian. The freedom to redefine it by any constant, however, allow us to use the following procedure. First of all, we define the *normal ordered product* of free fields as the operator obtained by expanding them in creation and annihilation operators, and reordering all the strings of creation and annihilation operators by placing all the creation operators to the left, and all the annihilation operators to the right, while leaving the corresponding coefficients untouched. In the case of two fields discussed here, for example, we have

$$\begin{aligned}
:\phi(x)\phi(y)^\dagger: \;&= \int d\Omega_p \int d\Omega_q :\left[a(\vec{p})e^{-ip\cdot x} + b(\vec{p})^\dagger e^{ip\cdot x}\right]\left[a(\vec{q})^\dagger e^{iq\cdot y} + b(\vec{p})e^{-iq\cdot y}\right]: \\
&\equiv \int d\Omega_p \int d\Omega_q \left[a(\vec{q})^\dagger a(\vec{p})e^{-ip\cdot x}e^{iq\cdot y} + a(\vec{p})b(\vec{p})e^{-ip\cdot x}e^{-iq\cdot y}\right. \\
&\qquad\qquad \left. +b(\vec{p})^\dagger a(\vec{q})^\dagger e^{ip\cdot x}e^{iq\cdot y} + b(\vec{p})^\dagger b(\vec{p})e^{ip\cdot x}e^{-iq\cdot y}\right]\,.
\end{aligned} \tag{6.59}$$

A more inteligible expression can be written if we define the *positive-frequency* and *negative-frequency* components of the field,

$$\begin{aligned} \phi_+(x) &\equiv \int d\Omega_p\, a(\vec{p}) e^{-ip\cdot x}\,, \qquad \phi_-(x) \equiv \int d\Omega_p\, b(\vec{p})^\dagger e^{ip\cdot x}\,, \\ \phi_+^*(x) &\equiv \int d\Omega_p\, b(\vec{p}) e^{-ip\cdot x}\,, \qquad \phi_-^*(x) \equiv \int d\Omega_p\, a(\vec{p})^\dagger e^{ip\cdot x}\,, \end{aligned} \tag{6.60}$$

so that $\phi(x) = \phi_+(x) + \phi_-(x)$ and $\phi(x)^\dagger = \phi_+^*(x) + \phi_-^*(x)$. In terms of these, the normal-ordered product is obtained by placing all the positive-frequency components on the right and all the negative-frequency components on the left,

$$:\phi(x)\phi(y)^\dagger: = \phi_+(x)\phi_+^*(x) + \phi_-^*(x)\phi_+(x) + \phi_-(x)\phi_+^*(x) + \phi_-(x)\phi_-^*(x)\,. \tag{6.61}$$

Whichever way we do it, it is clear that the normal-ordered product has zero vacuum expectation value. In the case of two fields, this is the only difference between the two type of products, i.e.,

$$\phi(x)\phi(y)^\dagger =: \phi(x)\phi(y)^\dagger : + [\phi_+(x), \phi_-^*(x)] =: \phi(x)\phi(y)^\dagger : + \langle 0|\phi(x)\phi(y)^\dagger|0\rangle\,. \tag{6.62}$$

This is particular instance of a general result known as Wick's theorem, that we will meet again later. This also means that the normal product of two fields obviously will not have the singularity we found in the usual product. We then give a more precise definition if the Hamiltonian using normal-ordering:

$$H = \int d^3x\, :\left\{\pi(x)^\dagger\pi(x) + [\vec{\nabla}\phi(x)]^\dagger\cdot[\vec{\nabla}\phi(x)] + m^2\phi(x)^\dagger\phi(x)\right\}:\,. \tag{6.63}$$

Now everything is well defined and, repeating the calculation, the only difference we find is

$$\begin{aligned} H &= \int d\Omega_p\, p^0 :\left[a(\vec{p})^\dagger a(\vec{p}) + b(\vec{p})b(\vec{p})^\dagger\right]: \\ &= \int d\Omega_p\, p^0 \left[a(\vec{p})^\dagger a(\vec{p}) + b(\vec{p})^\dagger b(\vec{p})\right]\,. \end{aligned} \tag{6.64}$$

This is the Hamiltonian of a system of scalar particles, and this should convince that we are on the right track.

6.5 Lagrangian formalism

The main drawback of the Hamiltonian formalism is that, since it singles out time with respect to space, is not manifestly relativistically-invariant. A manfestly invariant formalism is the Lagrangian formalism, and we know how to pass from one formalism to the other. Indeed, given conjugate canonical variables q and p and a Hamiltonian H, the Lagrangian L is obtained through

$$L(q,\dot{q}) = \dot{q}p - H(q,p)\,, \tag{6.65}$$

where we have to substitute $p = p(q,\dot{q})$ after finding $\dot{q} = \frac{\partial H}{\partial p}$. In the case at hand $\dot{\phi}(t,\vec{x}) = \pi(t,\vec{x})^\dagger$, so

$$\begin{aligned} L &= \int d^3x :\left\{2\dot{\phi}(x)^\dagger\dot{\phi}(x) - \dot{\phi}(x)^\dagger\dot{\phi}(x) - [\vec{\nabla}\phi(x)]^\dagger\cdot[\vec{\nabla}\phi(x)] - m^2\phi(x)^\dagger\phi(x)\right\}: \\ &= \int d^3x :\left\{\partial_\mu\phi(x)^\dagger\partial^\mu\phi(x) - m^2\phi(x)^\dagger\phi(x)\right\}:\,, \end{aligned} \tag{6.66}$$

where we have made use of normal ordering to avoid short-distance singularities.

At this point we can reverse the logic, and state a simple recipe to quantise the scalar field:

- take the following *Lagrangian density*

$$\mathscr{L}(x) = \ : \partial_\mu \phi(x)^\dagger \partial^\mu \phi(x) - m^2 \phi(x)^\dagger \phi(x) : ; \tag{6.67}$$

- construct the *action* integrating over spacetime,

$$S = \int d^4x \, \mathscr{L}(x) \, ; \tag{6.68}$$

- derive the equations of motion via the variational principle, $\delta S = 0$;
- impose the *canonical commutation relations* among the fields and their conjugate momenta.

What we have just outlined is called *canonical quantisation*, which we now describe in more detial.

One starts with a Lagrangian density, which is in general some real function of a set of fields and their derivatives, $\mathscr{L} = \mathscr{L}(\phi_i(x), \partial_\mu \phi_i(x))$.[40] The equations of motion are determined by means of a variational principle. For a given spacetime region D, one defines the action functional as follows,

$$S_D[\phi] = \int_D d^4x \, \mathscr{L}(\phi_i(x), \partial_\mu \phi_i(x)) \, . \tag{6.69}$$

The dynamics of the field is such that, for any D, the variation of the action vanishes, $\delta S_D = 0$, under arbitrary infinitesimal variations $\delta\phi_i$ of the fields that vanish on the boundary ∂D of D. The equations of motion that follow from this variational principle are the *Euler–Lagrange equations*. Let us derive them. Take a domain D and consider the infinitesimal variation $\phi_i(x) \to \phi(x) + \delta\phi_i(x)$, completely arbitrary except for the request $\delta\phi(x) = 0$ on ∂D. The variation of the action reads

$$\delta S_D[\phi] = \int_D d^4x \, \left[\mathscr{L}(\phi_i(x) + \delta\phi_i(x), \partial_\mu \phi_i(x) + \partial_\mu \delta\phi_i(x)) - \mathscr{L}(\phi_i(x), \partial_\mu \phi_i(x)) \right] . \tag{6.70}$$

Retaining only the first order in the variation we find

$$\delta S_D[\phi] = \int_D d^4x \, \sum_i \left[\delta\phi_i(x) \frac{\partial \mathscr{L}}{\partial \phi_i}(\phi_i(x), \partial_\mu \phi_i(x)) + \partial_\mu \delta\phi_i(x) \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)}(\phi_i(x), \partial_\mu \phi_i(x)) \right] . \tag{6.71}$$

Making use of Gauss' theorem we can integrate by parts the second term, obtaining

$$\begin{aligned} \delta S_D[\phi] = & \int_D d^4x \, \sum_i \delta\phi_i(x) \left[\frac{\partial \mathscr{L}}{\partial \phi_i}(\phi_i(x), \partial_\mu \phi_i(x)) - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)}(\phi_i(x), \partial_\mu \phi_i(x)) \right] \\ & + \int_{\partial D} d\Sigma_\mu \, \delta\phi_i(x) \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)}(\phi_i(x), \partial_\mu \phi_i(x)) \, . \end{aligned} \tag{6.72}$$

[40] We could include an explicit dependence on x, but we will be interested in translation-invariant Lagrangian densities only.

The boundary term vanishes per our request, and if the variation of the action must vanish for arbitrary variations of the fields in the interior of the domain, then the quantity in square brackets in the first term must vanish for every x. We then obtain the Euler–Lagrange equations,

$$\frac{\partial \mathscr{L}}{\partial \phi_i} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)} = 0\,. \tag{6.73}$$

To obtain the quantised fields we must first solve these equations, and then impose the canonical commuation relations. To do this we have of course to derive the canonical momenta conjugated ot the fields, which are defined as follows,

$$\pi_i \equiv \frac{\partial \mathscr{L}}{\partial(\partial_0 \phi_i)}\,. \tag{6.74}$$

Once we have computed these quantities, we impose the *equal-time commutation relations*, which play the role of canonical commutation relations for the the canonical coordinates $\phi_i(t, \vec{x}$ and their conjugate momenta $\pi_i(t, \vec{x})$,

$$\begin{aligned}
[\phi_i(x), \phi_j(y)]_{\mathrm{ET}} &= [\phi_i(t, \vec{x}), \phi_j(t, \vec{y})] = 0\,, \\
[\pi_i(x), \pi_j(y)]_{\mathrm{ET}} &= [\pi_i(t, \vec{x}), \pi_j(t, \vec{y})] = 0\,, \\
[\phi_i(x), \pi_j(y)]_{\mathrm{ET}} &= [\phi_i(t, \vec{x}), \pi_j(t, \vec{y})] = i\delta_{ij}\delta^{(3)}(\vec{x} - \vec{y})\,.
\end{aligned} \tag{6.75}$$

We can now construct the Hamiltonian via the usual Legendre transform,

$$H = \int d^3x\, \mathscr{H}(\phi_i(x), \pi_i(x)) = \int d^3x \left\{ \left[\sum_i \partial_0 \phi_i(x) \pi_i(x) \right] - \mathscr{L}(\phi_i(x), \partial_\mu \phi_i(x)) \right\}, \tag{6.76}$$

where $\mathscr{H}$ is the *Hamiltonian density*, and it is understood that we have to solve for $\partial_0 \phi_i(x)$ as a function of ϕ_i and π_i. The Hamiltonian determines the temporal evolution of fields and momenta through the Heisenberg equaiton,

$$\dot{\phi}(t, \vec{x}\,) = i[H, \phi(t, \vec{x}\,)]\,, \qquad \dot{\pi}(t, \vec{x}\,) = i[H, \pi(t, \vec{x}\,)]\,, \tag{6.77}$$

which is solved by

$$\phi(t, \vec{x}\,) = e^{iHt} \phi(0, \vec{x}\,) e^{-iHt}\,, \qquad \pi(t, \vec{x}\,) = e^{iHt} \pi(0, \vec{x}\,) e^{-iHt}\,. \tag{6.78}$$

This completes the canonical quantisation program.[41]

We may ask again why bother in developing these fancy formalisms. The answer is again that we are interested in building an interacting theory, and the canonical quantisation program automatically provides a causal theory. Moreover, the Lagrangian formalism makes manifest the symmetries of the theory, and guarantees the covariance of the equations of motions. Furthermore, it guarantees the existence of conserved quantities associated to the symmetries of the theory, and shows us how to build them, via the celebrated Noether's theorem.

[41] One should also provide a representation of the algebra Eq. (6.75) on some Hilbert space. Also, appropriate regularisation procedures, like normal ordering, might have to be used to deal with short-distance singularities.

6.6 Noether's theorem

Let us consider a generic Lagrangian and the corresponding action functional,

$$S = \int_D d^4x\, \mathscr{L}(\phi_i(x), \partial_\mu \phi_i(x), x)\,, \tag{6.79}$$

where for generality we have included also a possible x-dependence. Consider a mapping of fields and coordinates, $\phi_i'(x') = \mathcal{M}_i(\phi_j(x), x)$ and $x' = X(x)$, for some transformation $(\mathcal{M}, X)$, which we can think of as mapping from an observer $\mathcal{O}$ to another observer $\mathcal{O}'$, and define

$$S' = \int_{D'} d^4x'\, \mathscr{L}(\phi_i'(x'), \partial_\mu' \phi_i'(x'), x')\,, \tag{6.80}$$

where the *same* Lagrangian density is used, but the transformed fields and the transformed domain $D' = X(D)$ appear. In general S and S' differ: if we were to describe the same system in different reference frames, we will need in general different Lagrangian densities.

We now assume that S' and S differ only by boundary terms, i.e.,

$$S' = S + \int_D d^4x\, \partial_\mu F^\mu(\phi_i(x), x)\,, \tag{6.81}$$

for some function F^μ. In this case, the fields ϕ_i and ϕ_i' obey the same equations of motion. The proof is simple. Let ϕ_i be a solution of the EOM, and consider an infinitesimal arbitrary variation $\delta\phi_i(x)$ of the fields that vanishes on ∂D. Correspondingly, the field $\phi_i'(x') = \mathcal{M}_i(\phi_j(x), x)$ changes by $\delta\phi_i'(x') = \mathcal{M}_i(\phi_j(x) + \delta\phi_j(x), x) - \mathcal{M}_i(\phi_j(x), x)$, which clearly vanishes on $\partial D'$. Then from the definition of S', the usual variational procedure, and Eq. (6.81), we find

$$\begin{aligned}
\delta S' &= \int_{D'} d^4x' \sum_k \delta\phi_k'(x') \left[\frac{\partial \mathscr{L}}{\partial \phi_k}(\phi_i'(x'), \partial_\mu' \phi_i(x')) - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)}(\phi_i'(x'), \partial_\mu' \phi_i'(x'))\right] \\
&+ \int_{\partial D'} d\Sigma_\mu' \sum_k \delta\phi_k'(x') \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)}(\phi_i'(x'), \partial_\mu' \phi_i'(x')) \\
&= \int_{D'} d^4x' \sum_k \delta\phi_k'(x') \left[\frac{\partial \mathscr{L}}{\partial \phi_k}(\phi_i'(x'), \partial_\mu' \phi_i(x')) - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)}(\phi_i(x), \partial_\mu \phi_i(x))\right] \\
&= \delta S + \int_{\partial D} d\Sigma_\mu \sum_k \delta_k\phi(x) \frac{\partial F^\mu}{\partial \phi_k}(\phi_i(x), x) = \delta S = 0\,.
\end{aligned} \tag{6.82}$$

Since $\delta\phi$ are arbitrary, so are $\delta\phi'$, and for this equation to be satisfied one needs ϕ_i' to obey the same equations of motion as ϕ_i does.

Eq. (6.81) has another interesting consequence. Consider a continuous family of transformations for which Eq. (6.81) holds, and take an infinitesimal such transformation,

$$\begin{aligned}
\phi_i'(x') &= \phi_i(x) + \delta\phi_i(x) = \phi_i(x) + \epsilon M_i(\phi, x)\,, \\
x'^\mu &= x^\mu + \delta x^\mu = x^\mu + \epsilon \mathcal{A}^\mu(x)\,.
\end{aligned} \tag{6.83}$$

For future utility, let us note that

$$\begin{aligned}
\phi_i'(x) &= \phi_i'(x' - \delta x) = \phi_i'(x') - \delta x^\mu \partial_\mu \phi_i(x) = \phi_i(x) + \delta\phi_i(x) - \delta x^\mu \partial_\mu \phi_i(x) \\
&\equiv \phi_i(x) + \overline{\delta\phi}_i(x) = \phi_i(x) + \epsilon[M_i(\phi, x) - \mathcal{A}^\mu(x)\partial_\mu \phi_i(x)]\,.
\end{aligned} \tag{6.84}$$

Let us now change variables in S' back to x,

$$S' = \int_D d^4x \left| \det_{\mu\nu} \frac{\partial x'^{\,\mu}}{\partial x^\nu} \right| \mathscr{L}(\phi'(x+\delta x), \partial_\mu \phi'(x+\delta x), x+\delta x)\,, \tag{6.85}$$

and then expand in ϵ. Notice that to lowest order

$$\left| \det_{\mu\nu} \frac{\partial x'^{\,\mu}}{\partial x^\nu} \right| = \left| \det_{\mu\nu} \left(\delta^\mu{}_\nu + \frac{\partial \delta x^\mu}{\partial x^\nu} \right) \right| = \left| 1 + \mathrm{tr} \left(\frac{\partial \delta x^\mu}{\partial x^\nu} \right) \right| = \left| 1 + \partial_\mu \delta x^\mu \right| = 1 + \partial_\mu \delta x^\mu\,. \tag{6.86}$$

We then find

$$\begin{aligned} S' = \int_D d^4x \Big\{ & \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) + \partial_\mu \left[\delta x^\mu \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) \right] \\ & + \sum_i \left[\overline{\delta\phi}_i(x) \frac{\partial}{\partial \phi_i} \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) + \partial_\mu (\overline{\delta\phi}_i(x)) \frac{\partial}{\partial(\partial_\mu \phi_i)} \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) \right] \Big\}\,. \end{aligned} \tag{6.87}$$

Using Eq. (6.81) and integrating by parts we find

$$\begin{aligned} 0 = \int_D d^4x \Big\{ & \partial_\mu \left[\delta x^\mu \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) + \sum_i \overline{\delta\phi}_i(x) \frac{\partial}{\partial(\partial_\mu \phi_i)} \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) - F^\mu(\phi, x) \right] \\ & + \sum_i \overline{\delta\phi}_i(x) \left[\frac{\partial}{\partial \phi_i} \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) - \partial_\mu \frac{\partial}{\partial(\partial_\mu \phi_i)} \mathscr{L}(\phi(x), \partial_\mu \phi(x), x) \right] \Big\}\,. \end{aligned} \tag{6.88}$$

Imposing now the EOM, and given the arbitrariness of D and ϵ, we conclude that the *Noether current* J^μ,

$$\begin{aligned} & J^\mu(\phi, \partial_\mu \phi, x) = \\ & \qquad \mathcal{A}^\mu \mathscr{L}(\phi, \partial_\mu \phi, x) + \sum_i [M_i(\phi, x) - \mathcal{A}^\nu \partial_\nu \phi_i] \frac{\partial}{\partial(\partial_\mu \phi_i)} \mathscr{L}(\phi, \partial_\mu \phi, x) - F^\mu(\phi, x)\,, \end{aligned} \tag{6.89}$$

is a conserved current,

$$\partial_\mu J^\mu = 0\,. \tag{6.90}$$

From the Noether current one can easily construct a conserved *Noether charge*,

$$Q = \int d^3x\, J^0(x)\,. \tag{6.91}$$

A simple calculation shows that

$$\begin{aligned} \frac{d}{dt} Q &= \int d^3x\, \partial_0 J^0(x) = -\int d^3x\, \partial_j J^j(x) = -\lim_{V\to\infty} \int_V d^3x\, \partial_j J^j(x) \\ &= -\lim_{V\to\infty} \int_{\partial V} dn_j\, J^j(x) = 0\,, \end{aligned} \tag{6.92}$$

assuming that the fields are sufficiently well behaved at infinity, so that the flux of J^j at infinity vanishes. Explicitly,

$$\begin{aligned} Q &= \int d^3x \left[\mathcal{A}^0 \mathscr{L} + \sum_i [M_i - \mathcal{A}^\nu \partial_\nu \phi_i] \frac{\partial}{\partial(\partial_0 \phi_i)} \mathscr{L} - F^0 \right] \\ &= \int d^3x \left[\mathcal{A}^0 \left(\mathscr{L} - \sum_i \partial_0 \phi_i \pi_i \right) + \sum_i [M_i - \mathcal{A}^j \partial_j \phi_i] \pi_i - F^0 \right] \\ &= \int d^3x \left[\sum_i [M_i - \mathcal{A}^j \partial_j \phi_i] \pi_i - \mathcal{A}^0 \mathscr{H} - F^0 \right] . \end{aligned} \tag{6.93}$$

Let us consider a few explicit examples.

Translations The simplest case is that of invariance under translations. We assume that no boundary term appears, which is the case if there is no explicit dependence of the Lagrangian density on the coordinates (the only case that we will be considering in practice). In this case the action is obviously invariant. There are four kinds of translations, corresponding to four $\mathcal{A}^\mu_{(\nu)} = \delta^\mu{}_\nu$, with $M_i = 0$. The components of the corresponding conserved currents form the *canonical energy-momentum tensor*:

$$\Theta^\mu{}_\nu = -J^\mu_{(\nu)} = \sum_i \partial_\nu \phi_i \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)} - \delta^\mu{}_\nu \mathscr{L} , \qquad \Theta^{\mu\nu} = \sum_i \partial^\nu \phi_i \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)} - \eta^{\mu\nu} \mathscr{L} . \tag{6.94}$$

The reason for this nomenclature is easy: the $\mu = 0$ components of this tensor represent the energy and momentum density of the system, from which energy and momentum are obtained via integration:

$$\begin{aligned} \int d^3x \, \Theta^{00} &= \int d^3x \sum_i \partial^0 \phi_i \pi_i - \mathscr{L} = \int d^3x \, \mathscr{H} = H = P^0 , \\ \int d^3x \, \Theta^{0j} &= \int d^3x \sum_i \partial^j \phi_i \pi_i = - \int d^3x \sum_i \partial_j \phi_i \pi_i = P^j . \end{aligned} \tag{6.95}$$

Lorentz transformations The next, and most interesting case is that of Lorentz transformations. The transformation law of coordinates under an infinitesimal transformation is

$$x'^\mu = x^\mu + \frac{1}{2} \omega_{\rho\sigma} M^{(\rho\sigma)\mu}{}_\nu x^\nu , \tag{6.96}$$

where the matrices $M^{(\rho\sigma)}$ are given by

$$M^{(\rho\sigma)\,\mu}{}_\nu = \eta^{\rho\mu} \delta^\sigma{}_\nu - \eta^{\sigma\mu} \delta^\rho{}_\nu . \tag{6.97}$$

A scalar field $\phi(x)$ is, by definition, left invariant by a Lorentz transformation,

$$\phi'_a(x') = \phi_a(x) , \tag{6.98}$$

Plugging Eqs. (6.96) and (6.98) in the general expression for the Noether's current, Eq. (6.89), we find the conserved current $J^{(\rho\sigma)\mu}$ associated to the transformation with only $\omega_{\rho\sigma} \neq 0$,

$$\begin{aligned}
\mathcal{M}^{\mu,\rho\sigma} \equiv J^{(\rho\sigma)\mu} &= M^{(\rho\sigma)\,\mu}{}_{\nu} x^{\nu} \mathscr{L} - M^{(\rho\sigma)\,\nu}{}_{\alpha} x^{\alpha} \partial_{\nu}\phi_a \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_a)} \\
&= x^{\rho}\left(\partial^{\sigma}\phi_a \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_a)} - \eta^{\mu\sigma}\mathscr{L}\right) - x^{\sigma}\left(\partial^{\rho}\phi_a \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_a)} - \eta^{\mu\rho}\mathscr{L}\right) \\
&= x^{\rho}\Theta^{\mu\sigma} - x^{\sigma}\Theta^{\mu\rho}\,,
\end{aligned} \tag{6.99}$$

where $\Theta^{\mu\nu}$ is the energy-momentum tensor. The conserved charges are

$$\int d^3x\, \mathcal{M}^{0,\rho\sigma} = \int d^3x\, \left\{x^{\rho}\Theta^{0\sigma} - x^{\sigma}\Theta^{0\rho}\right\}\,. \tag{6.100}$$

Internal symmetries Certain theories are symmetric under transformations that involve only the fields and not the spacetime coordinates. Such transformations, mixing the various fields, are called internal transformations, and the corresponding symmetry is an *internal symmetry*. The simplest such transformations are linear in the fields, i.e.,

$$\delta\phi_i(x) = \epsilon \sum_j K_{ij}\phi_j(x)\,. \tag{6.101}$$

The corresponding conserved current and charge are easily determined,

$$\begin{aligned}
J^{\mu}_{\rm int} &= \sum_{ij} \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)} K_{ij}\phi_j\,, \\
Q_{\rm int} &= \int d^3x \sum_{ij} \frac{\partial \mathscr{L}}{\partial(\partial_0 \phi_i)} K_{ij}\phi_j = \int d^3x \sum_{ij} \pi_i K_{ij}\phi_j\,.
\end{aligned} \tag{6.102}$$

The derivation above is done in the classical case, but it can be extended to the quantum case as long as one takes carefully into account operator-ordering problems and short-distance singularity. In the case of free fields it suffices to normal-order the operators. The conserved-charge operators constructed in this way are seen to generate the corresponding symmetry transformations on the fields. Indeed, since $Q(t) = Q(0)$ can be taken at any time since it is conserved, we have that

$$\begin{aligned}
[i\epsilon Q, \phi_i(x)] &= \int d^3y \Big[\Big\{\sum_i [M_k(\phi(y), y) - \mathcal{A}^j(y)\partial_j\phi_k(y)]\pi_k(y) \\
&\qquad\qquad - \mathcal{A}^0(y)\mathscr{H}(\phi(y), \pi(y)) - F^0(\phi(y), y)\Big\}, \phi_i(x)\Big]_{\rm ET} \\
&= i\epsilon \int d^3y \sum_k \left\{M_k(\phi(y), y) - \mathcal{A}^j(y)\partial_j\phi_k(y) - \mathcal{A}^0(y)\frac{\mathscr{H}(\phi(y), \pi(y))}{\partial\pi_k(y)}\right\} \\
&\qquad\qquad \times [\pi_k(y), \phi_i(x)]_{\rm ET} \\
&= \epsilon \int d^3y \sum_k \Big\{M_k(\phi(y), y) - \mathcal{A}^j(y)\partial_j\phi_k(y) - \mathcal{A}^0(y)\partial_0\phi_k(y)\Big\}\delta_{ik}\delta^{(3)}(\vec{x} - \vec{y}\,) \\
&= \epsilon[M_i(\phi(x), x) - \mathcal{A}^{\nu}(x)\partial_\nu\phi_i(x)] = \overline{\delta\phi}_i(x)\,,
\end{aligned} \tag{6.103}$$

and so Q generates the symmetry transformation on the fields. For example,

$$\begin{aligned}
[P_\mu, \phi(x)] &= -i\partial_\mu \phi(x)\,, \\
[J^{(\rho\sigma)}, \phi(x)] &= i\left(x^\rho \partial^\sigma - x^\sigma \partial^\rho\right)\phi(x)\,, \\
[Q_{\rm int}, \phi_i(x)] &= -iK_{ij}\phi_j(x)\,.
\end{aligned} \tag{6.104}$$

These operators are seen to be Hermitian (after the ordering problems are dealt with), as long as we take the Lagrangian to be Hermitian (starting from a real Lagrangian at the classic level). Therefore, in the quantum case, Noether's theorem entails the existence of Hermitian generators of the continuous symmetries of the Lagrangian. This leads to a unitary representation of the Poincaré and other symmetry groups, under which the quantum field operators transform as their classical counterpart. This guarantees that our quantum system exhibits the desired symmetry.

6.7 Interacting fields and the interaction picture

So far we have discussed only the case of free fields. This is of quite limited utility to describe the real world, where particles interact in various ways. The next task is then that of building a theory that describes interacting particles while complying with the requests of Poincaré invariance and locality. The use of local fields allows to easily keep track of these two issues, thanks to their locality and their simple transformation properties. In the framework of field theory, the method of *canonical quantisation* is a convenient approach to the problem of building interacting theories: being based on the Lagrangian formalism, it exhibits manifestly the symmetries of the theory, and the requirement of microcausality is satisfied automatically by imposing the canonical commutation or anticommutation relations. Moreover, Noether's theorem allows us to construct the symmetry generators as operators on the Hilbert space of the system. Unfortunately, it is almost never possible to complete the canonical quantisation program in practice: the EOM for interacting theories are usually nonlinear, and it is not known how to solve them. It is therefore necessary to find some approximation technique that allows us to obtain something useful without having to solve the theory exactly.

In many cases of practical interest, the Hamiltonian of the system can be split into a free part and an interacting part, $H = H_0 + V$. Here the free part H_0 is the Hamiltonian of some system which we know how to solve explicitly, for example one of the free-field Hamiltonians discussed in the previous sections. The interaction part V contains every other term appearing in the full Hamiltonian. The idea is that V can be looked at as a perturbation to the free Hamiltonian H_0, and its effect evaluated in successive steps. This is the *perturbative quantisation* approach, which we now discuss in detail.

Let $\phi(x)$ be the field that realises the canonical quantisation program, and $\pi(x)$ its conjugate momentum. These fields solve the equations of motion and obey the CCR, and therefore

$$\begin{aligned}
\phi(x) = \phi(t,\vec{x}\,) &= e^{iHt}\phi(0,\vec{x}\,)e^{-iHt}\,, \qquad \dot\phi(t,\vec{x}\,) = i[H,\phi(t,\vec{x}\,)]\,, \\
\pi(x) = \pi(t,\vec{x}\,) &= e^{iHt}\pi(0,\vec{x}\,)e^{-iHt}\,, \qquad \dot\pi(t,\vec{x}\,) = i[H,\pi(t,\vec{x}\,)]\,.
\end{aligned} \tag{6.105}$$

The full Hamiltonian is obtained from the Lagrangian as usual,

$$H = \int d^3x\,[\pi(t,\vec{x}\,)\dot\phi(t,\vec{x}\,) - \mathscr{L}(\phi(t,\vec{x}\,),\partial\phi(t,\vec{x}\,))]\,, \tag{6.106}$$

where it is understood that $\partial_0\phi$ has to be expressed as a function of ϕ and π. In the cases we will consider, the full Lagrangian can be written as a free Lagrangian plus an interaction term

that depends on the fields but not on their derivatives, $\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_I$, with $\mathscr{L}_I = \mathscr{L}_I(\phi)$. It then follows that as a function of fields and their derivatives, the canonical momenta in the full interacting theory satisfy

$$\pi(\phi, \partial\phi) = \frac{\partial \mathscr{L}}{\partial(\partial_0\phi)} = \frac{\partial \mathscr{L}_0}{\partial(\partial_0\phi)} = \pi_0(\phi, \partial\phi)\,, \tag{6.107}$$

i.e., they have the same functional form as the conjugate momentum π_0 of the free theory; by the same token, $\partial_0\phi$ in the full theory will be the same function of ϕ and π as in the free theory. From Eq. (6.106) we then find in this case

$$\begin{aligned} H = H[\phi,\pi] &= \int d^3x\,[\pi(t,\vec{x})\dot{\phi}(t,\vec{x}) - \mathscr{L}_0(\phi(t,\vec{x}),\partial\phi(t,\vec{x})) - \mathscr{L}_I(\phi(t,\vec{x}))] \\ &= H_0[\phi,\pi] - \int d^3x\,\mathscr{L}_I(\phi(t,\vec{x})) = H_0[\phi,\pi] + V[\phi]\,. \end{aligned} \tag{6.108}$$

In the cases of interest, H is time-independent. On the other hand, after splitting it into H_0 and V, these will be separately time dependent. Let us do the splitting at $t = 0$,

$$H[\phi(t,\vec{x}),\pi(t,\vec{x})] = H[\phi(0,\vec{x}),\pi(0,\vec{x})] = H_0[\phi(0,\vec{x}),\pi(0,\vec{x})] + V[\phi(0,\vec{x})]\,. \tag{6.109}$$

From now on, H_0 and V will be those obtained using the interacting fields and momenta at $t = 0$. Let us now define the fields in the interaction picture as fields evolving in time with the free Hamiltonian, and coinciding with the full interacting fields (in the Heisenberg picture) at $t = 0$,

$$\begin{aligned} \phi_{\rm in}(t,\vec{x}) &\equiv e^{iH_0t}\phi_{\rm in}(0,\vec{x})e^{-iH_0t}\,, \qquad \phi_{\rm in}(0,\vec{x}) = \phi(0,\vec{x})\,, \\ \pi_{\rm in}(t,\vec{x}) &\equiv e^{iH_0t}\pi_{\rm in}(0,\vec{x})e^{-iH_0t}\,, \qquad \pi_{\rm in}(0,\vec{x}) = \pi(0,\vec{x})\,. \end{aligned} \tag{6.110}$$

At $t = 0$ the fields in the interaction picture obey the CCR, and since their values at time t is obtained via a unitary transformation, they will obey the CCR at all times. Since they evolve in time with the free Hamiltonian and obey the CCR, they automatically obey the Hamilton equations of motion of the free theory:

$$\dot{\phi}_{\rm in} = i[H_0,\phi_{\rm in}] = \frac{\delta H_0}{\delta\pi_{\rm in}}\,, \qquad \dot{\pi}_{\rm in} = i[H_0,\pi_{\rm in}] = -\frac{\delta H_0}{\delta\phi_{\rm in}}\,. \tag{6.111}$$

The fields in the interaction picture are then nothing else but free fields, and we already know exactly what they look like. If, for example, H_0 is the free Hamiltonian for the charged scalar field, we will have that

$$\phi_{\rm in}(t,\vec{x}) = \int d\Omega_p \left\{ a(\vec{p})e^{-ip\cdot x} + b(\vec{p})^\dagger e^{ip\cdot x} \right\}\,, \qquad \pi_{\rm in}(t,\vec{x}) = \dot{\phi}_{\rm in}(t,\vec{x})\,, \tag{6.112}$$

where $a(\vec{p})$, $a(\vec{p})^\dagger$, $b(\vec{p})$, $b(\vec{p})^\dagger$, are annihilation and creation operators that satisfy the usual commutation relations. At this point, we define also the interaction Hamiltonian in the interaction picture, $V_I(t)$, as

$$V_I(t) \equiv e^{iH_0t}V[\phi_{\rm in}(0,\vec{x})]e^{-iH_0t} = V[\phi_{\rm in}(t,\vec{x})] = -\int d^3x\,\mathscr{L}_I(\phi_{\rm in}(t,\vec{x}))\,. \tag{6.113}$$

Although it seems that we have made progress, in practice it is so only marginally: for example, if we want to determine the spectrum of the theory, we still have to solve the same eigenvalue

problem $H|\psi\rangle = E|\psi\rangle$, and although we have expressed H as a functional of free fields, this does not make the eigenvalue problem any easier to solve. On the other hand, if we now assume that the interaction V is small, for example because it enters the Hamiltonian with some small numerical prefactor, then we can attack the eigenvalue problem perturbatively, and solve it by successive approximations.

6.8 Perturbation theory and Wick's theorem

As in the problem of finding the spectrum of the theory, it takes the exact solution of the theory to know exactly the S-operator. On the other hand, if V, and so V_I, is a small perturbation, then we can power-expand S,

$$\begin{aligned} S &= \mathrm{Texp}\left\{-i\int_{-\infty}^{+\infty} dt\, V_I(t)\right\} \\ &= \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{-\infty}^{+\infty} d\tau_1 \ldots \int_{-\infty}^{+\infty} d\tau_n T\left\{V_I(\tau_1)\ldots V_I(\tau_n)\right\}\,, \end{aligned} \tag{6.114}$$

and compute the S-matrix elements order by order in the perturbation. We already know that products of fields at the same spacetime point lead to problems with infinities. To get rid of (part of) these, we take V_I to be normal-ordered: this does not change the symmetries of the S operator. Recalling again Eq. (6.113), and imposing normal ordering, we have that

$$\begin{aligned} S &= \mathrm{Texp}\left\{i\int d^4x\, :\mathscr{L}_I(\phi_{\rm in}(x)):\right\} \\ &= \sum_{n=0}^{\infty}\frac{i^n}{n!}\int d^4x_1 \ldots \int d^4x_n\, T\left\{:\mathscr{L}_I(\phi_{\rm in}(x_1)):\ldots:\mathscr{L}_I(\phi_{\rm in}(x_n)):\right\}\,. \end{aligned} \tag{6.115}$$

This formula is the starting point for the perturbative calculation of scattering cross sections.

In experiments, the case of the final state coinciding with the initial state is practically impossible to observe, as it would interfere with the experimental setup. Even if it were possible to consider this case, it would be impossible to distinguish between the lack of any interaction between the colliding particles, and the presence of some interaction that however gives a final state coinciding with the initial one. For these reasons, what is studied in practice is the case of final and initial states being different. It is then customary to explicitly subtract the no-interaction contribution to the S-matrix. The basic assumption of perturbation theory is that the exact S-matrix elements $\langle\varphi_f|S|\varphi_i\rangle$ can be well approximated with the lowest-order terms of the expansion of S, Eqs. (6.114) and (6.115). Subtracting the non-interaction term we have[42]

$$\langle\varphi_f|S-\mathbf{1}|\varphi_i\rangle = \langle\varphi_f|i\int d^4x :\mathscr{L}_I(x): + \frac{i^2}{2}\int d^4x_1\int d^4x_2 T\left\{:\mathscr{L}_I(x_1)::\mathscr{L}_I(x_2):\right\} + \ldots|\varphi_i\rangle\,. \tag{6.116}$$

The basic object of interest is therefore the matrix element

$$\langle\varphi_f|\int d^4x_1\int d^4x_2\ldots\int d^4x_n T\left\{:\mathscr{L}_I(x_1)::\mathscr{L}_I(x_2):\ldots:\mathscr{L}_I(x_n):\right\}|\varphi_i\rangle\,. \tag{6.117}$$

[42] We adopt the simplified notation $\mathscr{L}_I(x) = \mathscr{L}_I(\phi_{\rm in}(x))$.

The initial and final states are typically taken to be eigenstates of energy and momentum. More precisely, $|\varphi_{i,f}\rangle$ are taken to be eigenstates of the free Hamiltonian H_0 and of the free spatial momentum operators, built out of the free fields ϕ_{in}. Initial and final particle states are therefore created out of the vacuum states $|0\rangle$ by the corresponding creation operators. It is worth noting that thanks to translation invariance, if we denote with P_i and P_f the total initial and final four momentum, we find

$$\begin{aligned}
&\langle\varphi_f| \int d^4x_1 \ldots \int d^4x_n T\left\{ : \mathscr{L}_I(x_1) : \ldots : \mathscr{L}_I(x_n) : \right\} |\varphi_i\rangle \\
&= \int d^4x_1 \ldots \int d^4x_n \, \langle\varphi_f| e^{-ix_n\cdot P} T\left\{ : \mathscr{L}_I(x_1 - x_n) : \ldots : \mathscr{L}_I(0) : \right\} e^{ix_n\cdot P} |\varphi_i\rangle \\
&= \int d^4x_n e^{-ix_n\cdot(P_f - P_i)} \\
&\quad \times \langle\varphi_f| \int d^4y_1 \ldots \int d^4y_{n-1} T\left\{ : \mathscr{L}_I(y_1) : \ldots : \mathscr{L}_I(y_{n-1}) :: \mathscr{L}_I(0) : \right\} |\varphi_i\rangle \\
&= (2\pi)^4 \delta^{(4)}(P_f - P_i) \\
&\quad \times \langle\varphi_f| \int d^4y_1 \ldots \int d^4y_{n-1} T\left\{ : \mathscr{L}_I(y_1) : \ldots : \mathscr{L}_I(y_{n-1}) :: \mathscr{L}_I(0) : \right\} |\varphi_i\rangle \,,
\end{aligned} \tag{6.118}$$

i.e., the momentum-conserving delta function discussed above.

The matrix elements of Eq. (6.117) can be reduced to simpler ones by making use of *Wick's theorem*. In the case of a single Hermitian scalar field, this theorem states the following:

$$\begin{aligned}
T\left(\varphi(x_1)\ldots\varphi(x_n)\right) = \sum_{m=0}^{\left[\frac{n}{2}\right]} \Big\{ &: \varphi(x_1)\ldots\varphi(x_{n-2m}) : D(x_{n-2m+1}, x_{n-2m+2}) \ldots D(x_{n-1}, x_n) \\
&+ \text{other pairings} \Big\} \,.
\end{aligned} \tag{6.119}$$

Here the sum is over the number m of pairings of coordinates $\{x_1, \ldots, x_n\}$, from 0 to the maximal possible value, i.e., the integer part $\left[\frac{n}{2}\right]$ of $\frac{n}{2}$, and over the all the possible such pairings. The quantity $D(x, y)$, which we will refer to as the *contraction* of two fields, or the *propagator*, is given by

$$\begin{aligned}
D(x,y) &= \langle 0| T\left(\varphi(x)\varphi(y)\right)|0\rangle = \langle 0| e^{-iy\cdot P} T\left(\varphi(x-y)\varphi(0)\right) e^{iy\cdot P} |0\rangle \\
&= \langle 0| T\left(\varphi(x-y)\varphi(0)\right)|0\rangle = D(x-y) \,.
\end{aligned} \tag{6.120}$$

Notice also that for a real field

$$D(x-y) = \langle 0| T\left(\varphi(x)\varphi(y)\right)|0\rangle = \langle 0| T\left(\varphi(y)\varphi(x)\right)|0\rangle = D(y-x) \,. \tag{6.121}$$

We will not give a complete proof of this theorem, and limit ourselves to a sketch of a proof. We begin with the simplest case, $n = 2$. In this case, denoting with $\varphi_\pm$ the positive and negative frequency parts of the field, we have

$$\begin{aligned}
\varphi(x_1)\varphi(x_2) &= [\varphi_+(x_1) + \varphi_-(x_1)][\varphi_+(x_2) + \varphi_-(x_2)] =: \varphi(x_1)\varphi(x_2) : + [\varphi_+(x_1), \varphi_-(x_2)] \\
&=: \varphi(x_1)\varphi(x_2) : + \langle 0|[\varphi_+(x_1), \varphi_-(x_2)]|0\rangle =: \varphi(x_1)\varphi(x_2) : + \langle 0|\varphi_+(x_1)\varphi_-(x_2)|0\rangle \\
&=: \varphi(x_1)\varphi(x_2) : + \langle 0|\varphi(x_1)\varphi(x_2)|0\rangle \,,
\end{aligned} \tag{6.122}$$

where we have made use of the fact that $[\varphi_+(x_1), \varphi_-(x_2)]$ is a c-number and is thus identical to its vacuum expectation value, and further modified its expression by including terms that annihilate the vacuum and therefore have no effect. Imposing now time ordering, since by the way the normal-ordered product is defines one has $:\varphi(x_1)\varphi(x_2):=:\varphi(x_2)\varphi(x_1):$, we find

$$T\left(\varphi(x_1)\varphi(x_2)\right) =: \varphi(x_1)\varphi(x_2) : + \langle 0|T\left(\varphi(x_1)\varphi(x_2)\right)|0\rangle =: \varphi(x_1)\varphi(x_2) : + D(x_1, x_2)\,. \quad (6.123)$$

The proof for general n proceeds by induction, assuming that the theorem is true for n fields, and showing that it works for $n+1$. Writing the product of fields in terms of their positive-frequency and negative-frequency parts, we obtain a linear combination of strings of these operators. The key point is that whenever we bring a misplaced negative-frequency component from the right of the string to its left, passing through a positive-frequency component, we have to pay by adding a term where the pair of operators is replaced by their contraction. Doing this carefully, one sees that all the terms in Eq. (6.119) show up.

The result above in Eq. (6.119) is easily generalised to any number of real or complex scalar fields. Denoting them with φ_a, the only modification to the formula above is that contractions have to be considered only for fields of the same type, since $[\varphi_a, \varphi_b] = 0$ and $[\varphi_a, \varphi_b^\dagger] = 0$ for $a \neq b$, and moreover only for φ_a and its adjoint (which is the field itself in the real case), since for complex fields $[\varphi_a, \varphi_a] = 0$ and $[\varphi_a, \varphi_a^\dagger] \neq 0$.

In the case of interest to us, the fields in the time ordered product are already partially normal ordered. In this case, contractions have to be considered only among fields belonging to different normal-ordered blocks. This is easily understood from our sketch of a proof: a contraction appears only when we try to bring a negative-frequency field to the left of a positive-frequency field, but this never occurs for fields in a normal-ordered block.

6.9 Feynman diagrams

Making use of Wick's theorem one can compute the S-matrix elements, since the matrix elements of normal-ordered products of fields are easily obtained: they are given by the products of the coefficients of the annihilation operators that destroy a particle in the initial state, and those of the creation operators that destroy a particle in the final state, summed over all the possible ways in which particles and creation/annihilation operators can be paired. There is a convenient graphic method that allows to easily keep track of the various terms coming out of the perturbative expansion due to Wick's theorem. This is the method of Feynman diagrams, which we now discuss.

Consider a theory of a Hermitian (real) scalar field, to which we add a polynomial interaction, e.g., $\mathscr{L}_I(\phi) = \frac{\lambda}{4!}\phi^4$. In the interaction picture we find

$$\begin{aligned} S &= \mathrm{Texp}\left\{\frac{i\lambda}{4!}\int d^4x\ :\phi_{\mathrm{in}}(x)^4:\right\} \\ &= \sum_{n=0}^{\infty}\frac{(i\lambda)^n}{(4!)^n n!}\int d^4x_1 \ldots \int d^4x_n\, T\left\{:\phi_{\mathrm{in}}(x_1)^4: \ldots :\phi_{\mathrm{in}}(x_n)^4:\right\}\,. \end{aligned} \quad (6.124)$$

The free field $\phi_{\mathrm{in}}(x)$ describes free neutral scalar particles. Consider now the elastic scattering of two such particles, and compute perturbatively the corresponding scattering amplitude. To lowest order,

$$\langle \vec{p}_1'\vec{p}_2'|S - \mathbf{1}|\vec{p}_1\vec{p}_2\rangle = \frac{i\lambda}{4!}\int d^4x\, \langle \vec{p}_1'\vec{p}_2'| :\phi_{\mathrm{in}}(x)^4: |\vec{p}_1\vec{p}_2\rangle + \mathcal{O}(\lambda^2)\,. \quad (6.125)$$

The matrix element in Eq. (6.125) is easy to compute. Expanding the fields in the normal-ordered product in creation and annihilation operators, the only contribution will come from the terms in which there appear as many annihilation operators as in the initial state, and as many creation operator as in the final state. Any other term has either to many annihilation operators on its right side, so necessarily annihilating the initial state, or too many creation operators on its left side, so necessarily annihilating the final state. The creation and annihilation operators can be picked from any of the four fields, so there will be many such terms; on the other hand, they will all give the same contribution, since the four fields in the interaction term are equivalent. The degeneracy factor is easily seen to be 4!: one can freely choose which field will contribute the annihilation operator $a(\vec{p}_1)$, destroying particle 1 in the initial state, and this gives a factor of 4; the annihilation operator $a(\vec{p}_2)$, destroying particle 2 in the initial state, can be freely taken from one of the other fields, yielding a factor of 3; the creation operator $a(\vec{p}_1')^\dagger$, destroying particle 1 in the final state can now be chosen out of 2 fields, and $a(\vec{p}_1')^\dagger$ destroying particle 1 in the final state necessarily comes from the remaining field. We are then left with

$$\begin{aligned}&\langle \vec{p}_1'\vec{p}_2'|S-\mathbf{1}|\vec{p}_1\vec{p}_2\rangle \\ &= i\lambda\int d^4x\,\langle\vec{p}_1'|\phi_{\rm in}(x)|0\rangle\langle\vec{p}_2'|\phi_{\rm in}(x)|0\rangle\langle 0|\phi_{\rm in}(x)|\vec{p}_1\rangle\langle 0|\phi_{\rm in}(x)|\vec{p}_2\rangle+\mathcal{O}(\lambda^2)\,.\end{aligned} \tag{6.126}$$

The remaining matrix elements are easy to compute:

$$\begin{aligned}\langle 0|\phi_{\rm in}(x)|\vec{p}\rangle &= \int d\Omega_q\, e^{-iq\cdot x}\langle 0|a(\vec{q})|\vec{p}\rangle = \int d\Omega_q\, e^{-iq\cdot x}\langle 0|a(\vec{q})a(\vec{p})^\dagger|0\rangle \\ &= \int d\Omega_q\, e^{-iq\cdot x}\langle 0|[a(\vec{q}),a(\vec{p})^\dagger]|0\rangle = e^{-ip\cdot x}\,, \\ \langle \vec{p}|\phi_{\rm in}(x)|0\rangle &= \int d\Omega_q\, e^{iq\cdot x}\langle \vec{p}|a(\vec{q})^\dagger|0\rangle = \int d\Omega_q\, e^{iq\cdot x}\langle 0|a(\vec{p})a(\vec{q})^\dagger|0\rangle \\ &= \int d\Omega_q\, e^{iq\cdot x}\langle 0|[a(\vec{p}),a(\vec{q})^\dagger]|0\rangle = e^{ip\cdot x}\,.\end{aligned} \tag{6.127}$$

We then find

$$\begin{aligned}&\langle \vec{p}_1'\vec{p}_2'|S-\mathbf{1}|\vec{p}_1\vec{p}_2\rangle \equiv i(2\pi)^4\delta^{(4)}(p_1'+p_2'-p_1-p_2)\mathcal{M}(\vec{p}_1,\vec{p}_2;\vec{p}_1',\vec{p}_2') \\ &= i\lambda\int d^4x\, e^{i(p_1'+p_2'-p_1-p_2)\cdot x}+\mathcal{O}(\lambda^2) = i\lambda(2\pi)^4\delta^{(4)}(p_1'+p_2'-p_1-p_2)+\mathcal{O}(\lambda^2)\,.\end{aligned} \tag{6.128}$$

To order λ,

$$\mathcal{M}(\vec{p}_1,\vec{p}_2;\vec{p}_1',\vec{p}_2') = \lambda\,. \tag{6.129}$$

This is the quantity that enters the cross section.

This example was quite trivial, but nonetheless it shows already an important technical aspect, namely the need to properly count how many way there are to pair fields in the interaction Lagrangian and particles in the initial and final states. Here all the fields had to be paired with particles, but in more general cases we have to use Wick's theorem, and pair fields with other fields. Each factor of $\mathscr{L}_I$ in the perturbative expansion is called a *vertex*, and in the ϕ^4-theory described above each vertex comes with a factor of $i\lambda$.

Consider now a different type of interaction, like $\mathscr{L}_I(\phi)=\frac{\lambda}{3!}\phi^3$, and the same elastic scattering process described above. The $\mathcal{O}(\lambda)$ term in the perturbative expansion does not have

enough fields, and we need to go at least to the next perturbative order,

$$\langle \vec{p}_1' \vec{p}_2' | S - \mathbf{1} | \vec{p}_1 \vec{p}_2 \rangle = \frac{1}{2} \left(\frac{i\lambda}{3!} \right)^2 \int d^4x \int d^4y \, \langle \vec{p}_1' \vec{p}_2' | T \left\{ : \phi_{\rm in}(x)^3 :: \phi_{\rm in}(y)^3 : \right\} | \vec{p}_1 \vec{p}_2 \rangle + \mathcal{O}(\lambda^3) \, . \tag{6.130}$$

In order to compute the matrix element we now make use of Wick's theorem, and write the time-ordered product as a sum of normal-ordered products times contractions. We do not have to write the whole expansion, since only a normal-ordered product with four fields will have a non-vanishing matrix element. The term with four normal-ordered fields comes with one contraction, which can be done by picking any one of the three fields from each vertex, but not by picking two fields with the same vertex. There are therefore nine identical terms, and

$$\langle \vec{p}_1' \vec{p}_2' | S - \mathbf{1} | \vec{p}_1 \vec{p}_2 \rangle = \frac{1}{2} \left(\frac{i\lambda}{3!} \right)^2 3^2 \int d^4x \int d^4y \, D(x-y) \langle \vec{p}_1' \vec{p}_2' | : \phi_{\rm in}(x)^2 \phi_{\rm in}(y)^2 : | \vec{p}_1 \vec{p}_2 \rangle + \mathcal{O}(\lambda^3) \, . \tag{6.131}$$

The remaining matrix element is obtained by expanding the fields in creation and annihilation operator, and keeping only those terms with two creation operators and two annihilation operators. This can be done in different ways: we can take the two annihilation operators from fields at the same spacetime point, i.e., both from the two $\phi_{\rm in}(y)$ or both from the two $\phi_{\rm in}(x)$, or we can take one from one of the $\phi_{\rm in}(y)$ and one from one of the $\phi_{\rm in}(x)$. Let us begin with the first case, and count how many contributions are there: we can annihilate particle 1 in the initial state with an annihilation operator from either of the two $\phi_{\rm in}(y)$, or from either of the two $\phi_{\rm in}(x)$. Once we have chosen whether to use fields at y or fields at x for the initial state, we can still choose to take the creation operator that destroys particle 1 in the final state from either of the other two fields. Since we are integrating over x and y, they are equivalent and the same contribution is obtained after integration. All in all we have a factor of $2 \times 2 \times 2$ multiplying the contribution

$$\langle \vec{p}_1' | \phi_{\rm in}(x) | 0 \rangle \langle \vec{p}_2' | \phi_{\rm in}(x) | 0 \rangle \langle 0 | \phi_{\rm in}(y) | \vec{p}_1 \rangle \langle 0 | \phi_{\rm in}(y) | \vec{p}_2 \rangle = e^{i(p_1'+p_2')\cdot x} e^{-i(p_1+p_2)\cdot y} \tag{6.132}$$

to the matrix element in Eq. (6.131). If we instead take the annihilation operators from fields at different spacetime points, we can take the annihilation operator for particle 1 in the initial state from either of the two fields at x, or from either of the two fields at y; once this is done, we have to take the annihilation operator for particle 2 in the initial state from one of the two fields at the other spacetime point. This brings about a factor of $2 \times 2 \times 2$, since the two vertices are equivalent. We still have to choose which of the remaining two fields, living at different spacetime points, will provide the creation operator to annihilate particle 1 in the final state: this can be the same point from which the annihilation operator for particle 1 came, or that from which the annihilation operator for particle 2 came. In the first case we get the contribution

$$\langle \vec{p}_1' | \phi_{\rm in}(x) | 0 \rangle \langle \vec{p}_2' | \phi_{\rm in}(y) | 0 \rangle \langle 0 | \phi_{\rm in}(x) | \vec{p}_1 \rangle \langle 0 | \phi_{\rm in}(y) | \vec{p}_2 \rangle = e^{i(p_1'-p_1)\cdot x} e^{i(p_2'-p_2)\cdot y} \, , \tag{6.133}$$

which, as we said above, comes with a factor of $2 \times 2 \times 2$. In the other case we get the contribution

$$\langle \vec{p}_1' | \phi_{\rm in}(y) | 0 \rangle \langle \vec{p}_2' | \phi_{\rm in}(x) | 0 \rangle \langle 0 | \phi_{\rm in}(x) | \vec{p}_1 \rangle \langle 0 | \phi_{\rm in}(y) | \vec{p}_2 \rangle = e^{i(p_2'-p_1)\cdot x} e^{i(p_1'-p_2)\cdot y} \, , \tag{6.134}$$

again with a factor of $2 \times 2 \times 2$. Putting everything together, we find

$$\begin{aligned}\langle \vec{p}_1' \vec{p}_2' | S - \mathbf{1} | \vec{p}_1 \vec{p}_2 \rangle &= \frac{1}{2} \left(\frac{i\lambda}{3!} \right)^2 (3!)^2 2 \int d^4x \int d^4y \, D(x-y) \Big\{ e^{i(p_1'+p_2')\cdot x} e^{-i(p_1+p_2)\cdot y} \\ &+ e^{i(p_1'-p_1)\cdot x} e^{i(p_2'-p_2)\cdot y} + e^{i(p_2'-p_1)\cdot x} e^{i(p_1'-p_2)\cdot y} \Big\} + \mathcal{O}(\lambda^3) \\ &= (i\lambda)^2 \int d^4x \int d^4y \, D(x-y) \Big\{ e^{i(p_1'+p_2')\cdot x} e^{-i(p_1+p_2)\cdot y} \\ &+ e^{i(p_1'-p_1)\cdot x} e^{i(p_2'-p_2)\cdot y} + e^{i(p_2'-p_1)\cdot x} e^{i(p_1'-p_2)\cdot y} \Big\} + \mathcal{O}(\lambda^3) \, . \end{aligned} \tag{6.135}$$

To proceed in the calculation, let us write

$$D(x) = \int \frac{d^4q}{(2\pi)^4} \, e^{-iq\cdot x} \tilde{D}(q) \, , \tag{6.136}$$

to find

$$\begin{aligned}\langle \vec{p}_1' \vec{p}_2' | S - \mathbf{1} | \vec{p}_1 \vec{p}_2 \rangle &= (i\lambda)^2 \int d^4x \int d^4y \int \frac{d^4q}{(2\pi)^4} \, \tilde{D}(q) e^{-iq\cdot(x-y)} \Big\{ e^{i(p_1'+p_2')\cdot x} e^{-i(p_1+p_2)\cdot y} \\ &+ e^{i(p_1'-p_1)\cdot x} e^{i(p_2'-p_2)\cdot y} + e^{i(p_2'-p_1)\cdot x} e^{i(p_1'-p_2)\cdot y} \Big\} + \mathcal{O}(\lambda^3) \\ &= (i\lambda)^2 \int d^4x \int d^4y \int \frac{d^4q}{(2\pi)^4} \, \tilde{D}(q) \Big\{ e^{i(p_1'+p_2'-q)\cdot x} e^{-i(p_1+p_2-q)\cdot y} \\ &+ e^{i(p_1'-p_1-q)\cdot x} e^{i(p_2'-p_2+q)\cdot y} + e^{i(p_2'-p_1-q)\cdot x} e^{i(p_1'-p_2+q)\cdot y} \Big\} + \mathcal{O}(\lambda^3) \, . \end{aligned} \tag{6.137}$$

The integration over x and y can be carried out easily, and we get

$$\begin{aligned}&\langle \vec{p}_1' \vec{p}_2' | S - \mathbf{1} | \vec{p}_1 \vec{p}_2 \rangle \\ &= (i\lambda)^2 \int \frac{d^4q}{(2\pi)^4} \, \tilde{D}(q) \Big\{ (2\pi)^4 \delta^{(4)}(p_1' + p_2' - q)(2\pi)^4 \delta^{(4)}(p_1 + p_2 - q) \\ &\qquad + (2\pi)^4 \delta^{(4)}(p_1' - p_1 - q)(2\pi)^4 \delta^{(4)}(p_2' - p_2 + q) \\ &\qquad + (2\pi)^4 \delta^{(4)}(p_2' - p_1 - q)(2\pi)^4 \delta^{(4)}(p_1' - p_2 + q) \Big\} + \mathcal{O}(\lambda^3) \\ &= i(2\pi)^4 \delta^{(4)}(p_1' + p_2' - p_1 - p_2) \Big\{ i\lambda^2 \Big[\tilde{D}(p_1 + p_2) + \tilde{D}(p_1' - p_1) + \tilde{D}(p_2' - p_1) \Big] \Big\} \, . \end{aligned} \tag{6.138}$$

The quantity in braces is the matrix element $\mathcal{M}_{fi}$ entering the cross section. This calculation taught us another couple of points, namely the appearance of momentum-conserving delta functions at each vertex once that integration over the vertex position has been carried out, and that calculations get easier in momentum space. We also should be confident enough with identifying the different contribution to the scattering amplitude obtained by contracting fields either with other fields (from a different vertex) or with initial or final-state particles, and with counting how many times each contribution appears.

A convenient way to obtain the result above with much less effort is via a graphic device known as *Feynman diagram*. The procedure is as follows. For a given set of initial and final-state particles, and at a given perturbative order:

- draw a point with as many lines coming out of it as fields in an interaction vertex (in general one should distinguish between different kinds of fields, and between fields and their adjoints: this is immaterial here);
- associate all the incoming or outgoing particles with one of the lines/fields, and pair all the remaining lines/fields with each other (in general only certain fields can be associated with an external particle of a certain type, but here there is only one neutral particle; for the same reason, while in general each lines is given an orientation, for neutral scalar particles this does not matter; if a particle or a field always remains unpaired, discard the graph and go over to the next perturbative order);
- draw all the possible *topologically inequivalent* diagrams with the prescribed external particles and number of vertices, i.e., diagrams that cannot be deformed into one another without cutting or glueing parts of them;
- count how many times a given diagram is obtained by exchanging equivalent lines or vertices;
- associate a momentum to each line of the graph: if it is a line ending in one of the initial or final state particles, let the corresponding momentum flow into the graph for initial-state particles, and out of the graph for final-state particles; if it is an internal line, choose the direction in which the momentum flows arbitrarily;
- for each external line ending in a particle, include a factor of 1 (it is trivial in this case, but for particles other than scalars there are nontrivial factors);
- for each internal line include a factor $\tilde{D}(q)$, with q the momentum associated to that line;
- for each vertex, write down a factor $(2\pi)^4\delta^{(4)}(\sum_j p_j)$, where p_j are the momenta flowing *into* the vertex (this enforces momentum conservation at each vertex);
- integrate over all internal momenta.

The counting part can be tricky sometimes, and it is better done by choosing a certain topology, which in the case at hand amounts to the connectivity properties of the various vertices, and counting in how many ways that is achieved by associating fields with the initial and final-state particles and with other fields.

With these simple rules is a matter of seconds to obtain Eq. (6.138). The only bit that remains to be done is computing $\tilde{D}(q)$. In coordinate space it is straightforward to get

$$\begin{aligned} D(x) &= \langle 0|T\left(\phi(x)\phi(y)\right)|0\rangle = \langle 0|\theta(x^0-y^0)\phi(x)\phi(y)+\theta(y^0-x^0)\phi(y)\phi(x)|0\rangle \\ &= \int d\Omega_p \int d\Omega_q \langle 0|a(\vec{p})a(\vec{q})^\dagger|0\rangle \left[\theta(x^0-y^0)e^{-ip\cdot x}e^{iq\cdot y}+\theta(y^0-x^0)e^{-ip\cdot y}e^{iq\cdot x}\right] \\ &= \int d\Omega_p \left[\theta(x^0-y^0)e^{-ip\cdot(x-y)}+\theta(y^0-x^0)e^{-ip\cdot(y-x)}\right]. \end{aligned} \tag{6.139}$$

In order to obtain the Fourier transform of this quantity, it is convenient to notice that $D(x)$ obey the inhomogeneous Klein-Gordon equation. Indeed, we have

$$\begin{aligned} &\partial^2_{x0}T\left(\phi(x)\phi(y)\right) = \partial_{x0}\left\{T\left(\partial_{x0}\phi(x)\phi(y)\right)+\delta(x^0-y^0)[\phi(x),\phi(y)]_{\rm ET}\right\} \\ &= \partial_{x0}T\left(\partial_{x0}\phi(x)\phi(y)\right) = T\left(\partial^2_{x0}\phi(x)\phi(y)\right)+\delta(x^0-y^0)[\partial_{x0}\phi(x),\phi(y)]_{\rm ET} \\ &= T\left(\partial^2_{x0}\phi(x)\phi(y)\right)+\delta(x^0-y^0)[\pi(x),\phi(y)]_{\rm ET} = T\left(\partial^2_{x0}\phi(x)\phi(y)\right)-i\delta^{(4)}(x-y), \end{aligned} \tag{6.140}$$

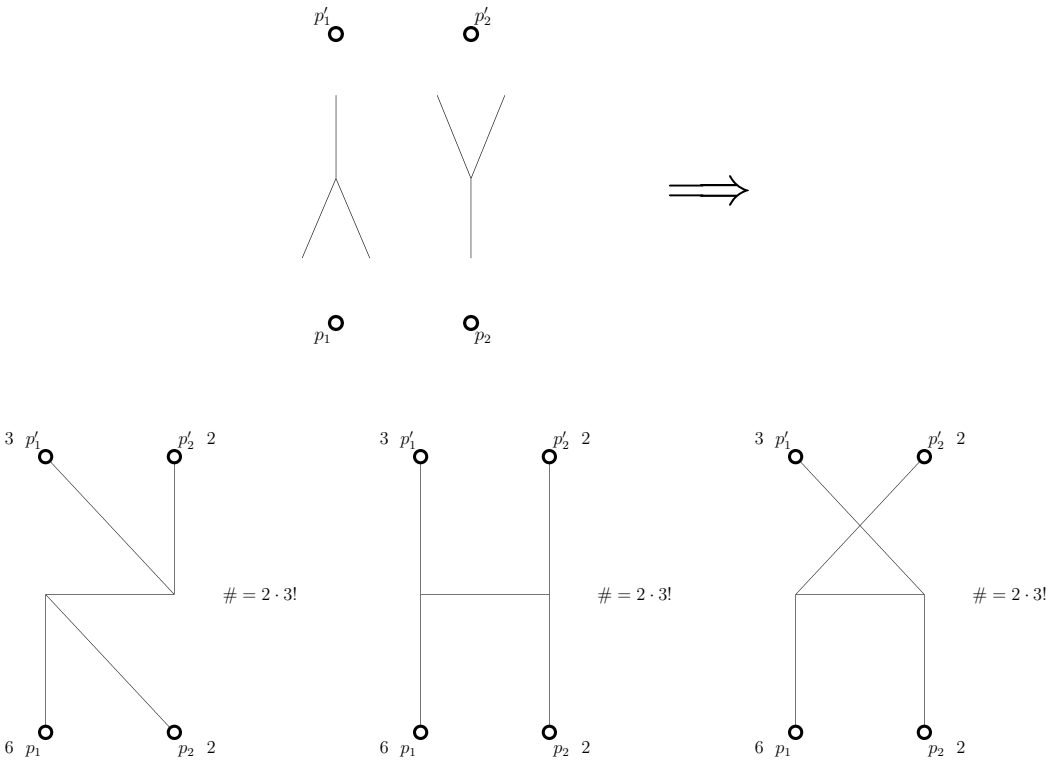

Figure 30: Topologically distinct diagrams for $2 \to 2$ elastic scattering at the lowest perturbative order in the ϕ^3 theory, obtained connecting fields from two vertices and four external particles. The number of times each topologically distinct configuration is also reported.

and so we find

$$(\Box_x + m^2)D(x-y) = \langle 0|T\left((\Box_x + m^2)\phi(x)\phi(y)\right) - i\delta^{(4)}(x-y)|0\rangle = -i\delta^{(4)}(x-y)\,. \quad (6.141)$$

Going over to momentum space we find immediately that $\tilde{D}(q) = i/(q^2 - m^2)$, except near the singularity at $q^2 = m^2$. This singularity must be dealt with by choosing some prescription when integrating $\tilde{D}(q)$ over momentum, and this must be done so that Eq. (6.139) is reproduced. The correct prescription is

$$\tilde{D}(p) = \frac{i}{p^2 - m^2 + i\epsilon}\,. \quad (6.142)$$

Indeed, the poles of this expression are at $p^0 = \pm\sqrt{\vec{p}^2 + m^2} \mp i\epsilon$, and so using the residue theorem we find

$$\begin{aligned}
&\int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x}\tilde{D}(p) = \int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x}\frac{i}{p^2 - m^2 + i\epsilon} \\
&= i\int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x}\frac{1}{p^0 - \sqrt{\vec{p}^2 + m^2} + i\epsilon}\frac{1}{p^0 + \sqrt{\vec{p}^2 + m^2} - i\epsilon} \\
&= i\int \frac{d^3p}{(2\pi)^4}\left\{\theta(x^0)(-2\pi i)e^{-ip\cdot x}\frac{1}{2p^0} + \theta(-x^0)(2\pi i)e^{-ip\cdot x}\frac{1}{-2p^0}\right\} \\
&= i\int \frac{d^3p}{(2\pi)^3 2p^0}\left\{\theta(x^0)e^{-ip\cdot x} + \theta(-x^0)e^{-ip\cdot x}\right\} = D(x)\,.
\end{aligned} \quad (6.143)$$

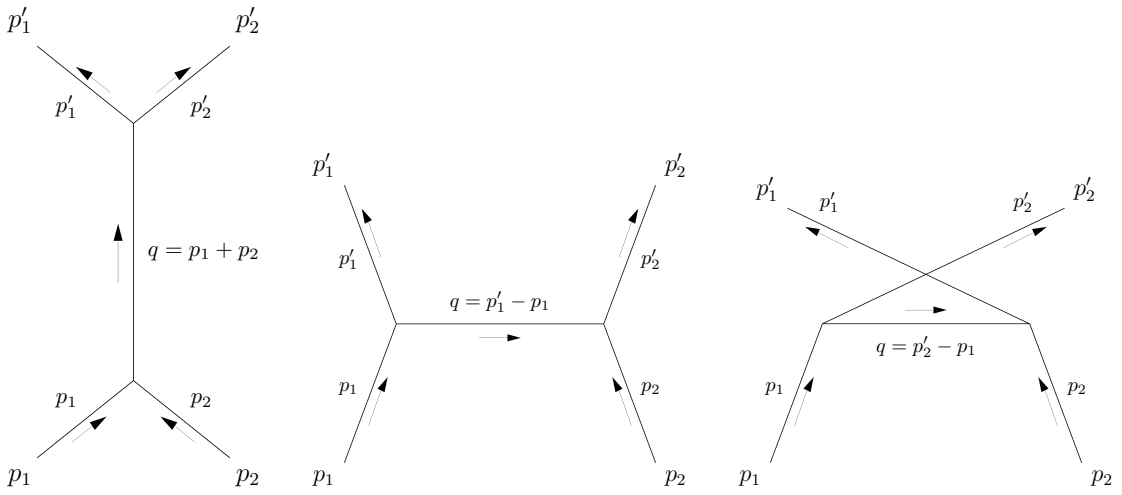

Figure 31: Feynman diagrams for $2 \to 2$ elastic scattering at the lowest perturbative order in the ϕ^3 theory. The momenta associated to external and internal lines are also shown.

We then have for $2 \to 2$ elastic scattering in the ϕ^3 theory

$$\begin{aligned} \mathcal{M}_{fi} &= i\lambda^2 \left\{ \frac{i}{(p_1+p_2)^2 - m^2} + \frac{i}{(p_1'-p_1)^2 - m^2} + \frac{i}{(p_2'-p_2)^2 - m^2} \right\} \\ &= -\lambda^2 \left\{ \frac{1}{s-m^2} + \frac{1}{t-m^2} + \frac{1}{u-m^2} \right\}, \end{aligned} \tag{6.144}$$

which manifestly displays crossing symmetry.

6.10 Fermions

So far we have discussed only scalar fields, but all the matter particles (quarks and leptons) are spin-$\frac{1}{2}$ fermions. This requires that we extend our formalism to include them.

Generalising Fock space is straightforward: it suffices to take into account that there is now one more quantum number for free fermions, namely the spin in one of the three spatial direction, which is usually taken to be the z direction. One-particle states therefore read

$$|\vec{p}\,, s_z\rangle \tag{6.145}$$

with $s_z = \pm\frac{1}{2}$. For multiparticle states one has to take into account the Fermi-Dirac statistics of fermion states, i.e., a the state of a system of many identical fermions must change sign under exchange of any two of them. We then have for multiparticle states

$$|\vec{p}_1\,, s_{1\,z}; \vec{p}_2\,, s_{2\,z}; \ldots \vec{p}_n\,, s_{n\,z};\rangle = (-1)^{\sigma_{\mathrm{P}}} |\vec{p}_{\mathrm{P}(1)}\,, s_{\mathrm{P}(1)\,z}; \vec{p}_{\mathrm{P}(2)}\,, s_{\mathrm{P}(2)\,z}; \ldots \vec{p}_{\mathrm{P}(n)}\,, s_{\mathrm{P}(n)\,z}\rangle \tag{6.146}$$

where P is any permutations of $\{1, 2, \ldots, n\}$, and σ_{P} is its signature, which equals 0 if the permutation is obtained with an even number of transpositions (exchanges of two elements), and 1 if the permutation is obtained with an odd number of transpositions. For example,

$$\begin{aligned} |\vec{p}_1\,, s_{1\,z}; \vec{p}_2\,, s_{2\,z}\rangle &= -|\vec{p}_2\,, s_{2\,z}; \vec{p}_1\,, s_{1\,z}\rangle\,, \\ |\vec{p}_1\,, s_{1\,z}; \vec{p}_2\,, s_{2\,z}; \vec{p}_3\,, s_{3\,z}\rangle &= -|\vec{p}_2\,, s_{2\,z}; \vec{p}_1\,, s_{1\,z}; \vec{p}_3\,, s_{3\,z}\rangle = |\vec{p}_2\,, s_{2\,z}; \vec{p}_3\,, s_{3\,z}; \vec{p}_1\,, s_{1\,z}\rangle\,. \end{aligned} \tag{6.147}$$

We can define creation and annihilation operators like in the scalar case, adding an extra index for the spin of the particle that is created or destroyed. Due to the Fermi-Dirac statistics of the staes, Eq. (6.146), however, the annihilation operators $b_s(\vec{p})$ and the creation operators $b_s(\vec{p})^\dagger$, $s = \pm\frac{1}{2}$, must obey *anticommutation relations*,

$$\begin{aligned} &\{b_s(\vec{p}), b_t(\vec{q})^\dagger\} \equiv b_s(\vec{p}) b_t(\vec{q})^\dagger + b_t(\vec{q})^\dagger b_s(\vec{p}) = \delta_{st}(2\pi)^3 2p^0 \delta^{(3)}(\vec{p}-\vec{q})\,, \\ &\{b_s(\vec{p}), b_t(\vec{q})\} = \{b_s(\vec{p})^\dagger, b_t(\vec{q})^\dagger\} = 0 \end{aligned} \tag{6.148}$$

These operators create particle states out of the vacuum state $|0\rangle$, which is annihilated by all the $b_s(\vec{p})$,

$$b_s(\vec{p})|0\rangle = 0\,, \quad \forall s\,, \forall \vec{p}. \tag{6.149}$$

If different species of fermions are present, the corresponding creation and annihilation operators are taken to anticommute; if fermions and bosons are present, their respective creation and annihilation operators are taken to commute.

6.10.1 The Dirac equation

We now want to build quantum fields to describe the dynamics of these particles, following a procedure similar to the one used for scalar particles: solve a relativistic wave equation and then impose quantisation of the amplitudes of the normal modes, or equivalently imposing canonical commutation relations (once we have identified the appropriate Lagrangian). The Klein-Gordon equation is the trasposition into quantum-mechanical language of the energy-momentum relation $E^2 = \vec{p}^2 + m^2$, and so has to be obeyed by the fermion field; on the other hand, there is nothing in that equation that specifies the number of spin states of a fermion.

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 still within relativistic quantum mechanics. These problems, as mentioned above, where the existence of negative energy states, and the impossibility to find a covariant probability current giving a positive-definite probability density.[43] 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 first-order equation for the fermions. 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

$$b^\mu \partial_\mu \psi = a\psi\,, \tag{6.152}$$

satisfied by the wave function ψ, which implies automatically

$$b^\mu \partial_\mu b^\nu \partial_\nu \psi = a^2 \psi \Longrightarrow (-b^\mu \partial_\mu b^\nu \partial_\nu + a^2)\psi = 0\,, \tag{6.153}$$

[43] From the Klein-Gordon equation and its complex conjugate one finds

$$0 = \phi^*(\Box + m^2)\phi - \phi(\Box + m^2)\phi^* = \phi^* \partial_\mu \partial^\mu \phi - \phi \partial_\mu \partial^\mu \phi^* = \partial_\mu[\phi^* \partial^\mu \phi - (\partial^\mu \phi^*)\phi]\,, \tag{6.150}$$

and so the current

$$J^\mu = i\phi^* \overset{\leftrightarrow}{\partial^\mu} \phi \tag{6.151}$$

is conserved. We note in passing that this is the Noether current associated to the U(1) symmetry of the Klein-Gordon Lagrangian under $\phi \to e^{i\alpha}$ and $\phi^\dagger \to e^{-i\alpha}$. This is therefore a good candidate for a probability current, since it is a four-vector and it is conserved, but unfortunately its $\mu = 0$ component is not positive-definite.

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

$$\begin{aligned} \Box &= -b^\mu b^\nu \partial_\mu \partial_\nu = -\tfrac{1}{2}\{b^\mu, b^\nu\}\partial_\mu\partial_\nu \,, \\ m^2 &= a^2 \,. \end{aligned} \tag{6.154}$$

which will be satisfied if

$$\{b^\mu, b^\nu\} = -2\eta^{\mu\nu} \,, \qquad a^2 = m^2 \,. \tag{6.155}$$

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 $(b^\mu)^2 = -1$ for all μ, which is incompatible with $b^\mu b^\nu = 0$ for $\mu \neq \nu$. The simplest way to solve Eq. (6.155) is by means of 4×4 matrices, $b^\mu = i\gamma^\mu$, $a = m\mathbf{1}_4$, with

$$\gamma^0 = \begin{pmatrix} \mathbf{1}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & -\mathbf{1}_2 \end{pmatrix} \,, \qquad \gamma^j = \begin{pmatrix} \mathbf{0}_2 & \sigma_j \\ -\sigma_j & \mathbf{0}_2 \end{pmatrix} \,. \tag{6.156}$$

One can verify explicitly that

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \,. \tag{6.157}$$

The resulting equation,

$$(i\not{\partial} - m)\psi = 0 \,, \tag{6.158}$$

where $\not{\partial} \equiv \gamma^\mu \partial_\mu$, is the *Dirac equation*. This equation solved the problem of the non-positive probability density, but did not solve the problem of negative energies. In fact, from this equation one can derive the Dirac Hamiltonian as follows,

$$i\partial_0\psi = (m\gamma^0 - i\vec{\nabla}\cdot\gamma^0\vec{\gamma})\psi \equiv H_{\rm Dirac}\psi \,. \tag{6.159}$$

Energy eigenfunctions are easily found in the form of plane waves, $\psi = \psi_0 e^{-ip\cdot x}$, with ψ_0 satisfying

$$\gamma^0(m + \vec{p}\cdot\vec{\gamma})\psi_0 = p^0\psi_0 \to (p^0\gamma^0 - \vec{p}\cdot\vec{\gamma} - m)\psi_0 = (\not{p} - m)\psi_0 = 0 \,. \tag{6.160}$$

There are four solutions to Eq. (6.160), 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 $\psi_- = v(\vec{p})e^{ip\cdot x}$. Here u and v denote four-component *bispinors*. One finds

$$(\not{p} - m)u(\vec{p}) = 0 \,, \qquad (\not{p} + m)v(\vec{p}) = 0 \,. \tag{6.161}$$

Setting

$$u = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \,, \qquad v = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \tag{6.162}$$

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

$$\begin{aligned} 0 = (\not{p} - m)u &= \begin{pmatrix} (p^0 - m)\xi_1 - \vec{p}\cdot\vec{\sigma}\xi_2 \\ -(p^0 + m)\xi_2 + \vec{p}\cdot\vec{\sigma}\xi_1 \end{pmatrix} \,, \\ 0 = (\not{p} + m)v &= \begin{pmatrix} (p^0 + m)\eta_1 - \vec{p}\cdot\vec{\sigma}\eta_2 \\ -(p^0 - m)\eta_2 + \vec{p}\cdot\vec{\sigma}\eta_1 \end{pmatrix} \,. \end{aligned} \tag{6.163}$$

These are solved setting

$$\xi_2 = \frac{\vec{p}\cdot\vec{\sigma}}{p^0+m}\xi_1\,, \qquad \eta_1 = \frac{\vec{p}\cdot\vec{\sigma}}{p^0+m}\eta_2\,. \tag{6.164}$$

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

$$u_s(\vec{p}) = \sqrt{p^0+m}\begin{pmatrix}\chi_s\\ \frac{\vec{p}\cdot\vec{\sigma}}{p^0+m}\chi_s\end{pmatrix}\,, \qquad v_s(\vec{p}) = \sqrt{p^0+m}\begin{pmatrix}\frac{\vec{p}\cdot\vec{\sigma}}{p^0+m}\tilde{\chi}_s\\ \tilde{\chi}_s\end{pmatrix}\,, \tag{6.165}$$

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

$$\begin{aligned} &\bar{u}_{s'}(\vec{p})u_s(\vec{p}) = 2m\delta_{s's}\,, && \bar{v}_{s'}(\vec{p})v_s(\vec{p}) = -2m\delta_{s's}\,,\\ &\bar{u}_{s'}(\vec{p})v_s(\vec{p}) = 0 && \bar{v}_{s'}(\vec{p})u_s(\vec{p}) = 0\,, \end{aligned} \tag{6.166}$$

where

$$\bar{u}_s(\vec{p}) \equiv u_s(\vec{p})^\dagger\gamma^0\,, \qquad \bar{v}_s(\vec{p}) \equiv v_s(\vec{p})^\dagger\gamma^0\,. \tag{6.167}$$

Completeness of the solutions entails the relations

$$\sum_s u_s(\vec{p})\bar{u}_s(\vec{p}) = \not{p} + m\,, \qquad \sum_s v_s(\vec{p})\bar{v}_s(\vec{p}) = \not{p} - m\,. \tag{6.168}$$

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

$$u_s(\vec{p}) \to \sqrt{2m}\begin{pmatrix}\chi_s\\ 0\end{pmatrix}\,, \tag{6.169}$$

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

$$\chi_1 = \begin{pmatrix}1\\0\end{pmatrix} \qquad \chi_2 = \begin{pmatrix}0\\1\end{pmatrix}\,, \tag{6.170}$$

so that the $s = 1$ solution corresponds to $s_z = \frac{1}{2}$ and the $s = 2$ solution corresponds to $s_z = -\frac{1}{2}$. 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. In the low-energy limit we find for the negative-energy solution

$$v_s(\vec{p}) \to \sqrt{2m}\begin{pmatrix}0\\ \tilde{\chi}_s\end{pmatrix}\,. \tag{6.171}$$

Since the change of sign of the energy can be obtained by changing the direction of time, consistency requires that we change both the spatial momenta and the spin of the particle, and so we set[44]

$$\tilde{\chi}_1 = \begin{pmatrix}0\\1\end{pmatrix} \qquad \tilde{\chi}_2 = -\begin{pmatrix}1\\0\end{pmatrix}\,, \tag{6.172}$$

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

[44] Ignore the minus sign for the time being. If you cannot ignore it, the reason for it is similar to the reason why a minus sign appeared when relating antiquarks and isospin states: this time the symmetry group we are representing is the Lorentz group, and if we take the particle to transform according to a certain representation, then the antiparticle must transform according to the complex conjugate of that representation.

6.10.2 Quantisation of the Dirac field

We now want to preceed with the second quantisation of the Diurac wave function, promoting it to a quantum fields. To do so, we write the most general solution of Eq. (6.158) as a linear superposition of the positive-energy and negative-energy solutions, Eq. (6.165),

$$\psi(x) = \int d\Omega_p \sum_s \left\{ u_s(\vec{p}) b_s(\vec{p}) e^{-ip\cdot x} + v_s(\vec{p}) d_s(\vec{p})^\dagger e^{ip\cdot x} \right\} , \tag{6.173}$$

where for the time being $b_s(\vec{p})$ and $d_s(\vec{p})^\dagger$ denote ordinary complex numbers. Notice that ψ has an index $\alpha = 1\, ldots, 4$, like u_s and v_s do. This result could have been obtained directly by first Fourier-transforming to momentum space the Dirac equation,

$$\psi(x) = \int \frac{d^4p}{(2\pi)^4}\, e^{-ip\cdot x} \tilde{\psi}(p) \,, \qquad (\not{p} - m)\tilde{\psi}(p) = 0 \,. \tag{6.174}$$

Multiplying the Dirac equation in momentum space by $\not{p} + m$ we find

$$0 = (\not{p} + m)(\not{p} - m)\tilde{\psi}(p) = (\not{p}^2 - m^2)\tilde{\psi}(p) = (p^2 - m^2)\tilde{\psi}(p) \,, \tag{6.175}$$

so that

$$\begin{aligned} \tilde{\psi}(p) &= 2\pi\delta(p^2 - m^2)\hat{\psi}(p^0, \vec{p}) \\ &= \frac{2\pi}{2\varepsilon(\vec{p})} \left\{ \delta(p^0 - \varepsilon(\vec{p}))\hat{\psi}(\varepsilon(\vec{p}), \vec{p}) + \delta(p^0 + \varepsilon(\vec{p}))\hat{\psi}(-\varepsilon(\vec{p}), \vec{p}) \right\} \\ &= \frac{2\pi}{2\varepsilon(\vec{p})} \left\{ \delta(p^0 - \varepsilon(\vec{p}))B(\vec{p}) + \delta(p^0 + \varepsilon(\vec{p}))D(-\vec{p}) \right\} \,, \end{aligned} \tag{6.176}$$

with

$$(\not{p} - m)B(\vec{p}) = 0 \,, \qquad (\not{p} + m)D(\vec{p}) = 0 \,, \tag{6.177}$$

where it is understood that $p^0 = \varepsilon(\vec{p})$. These equations are solved by the expressions of Eq. (6.165),

$$B(\vec{p}) = \sum_s u_s(\vec{p}) b_s(\vec{p}) \,, \qquad D(\vec{p}) = \sum_s v_s(\vec{p}) d_s(\vec{p})^\dagger \,, \tag{6.178}$$

for arbitrary complex $b_s(\vec{p})$ and $d_s(\vec{p})^\dagger$, and so

$$\begin{aligned} \psi(x) &= \int \frac{d^4p}{(2\pi)^4} \sum_s e^{-ip\cdot x} \frac{2\pi}{2\varepsilon(\vec{p})} \left\{ \delta(p^0 - \varepsilon(\vec{p}))u_s(\vec{p})b_s(\vec{p}) + \delta(p^0 + \varepsilon(\vec{p}))v_s(-\vec{p})d_s(-\vec{p})^\dagger \right\} \\ &= \int \frac{d^3p}{(2\pi)^3 2p^0} \sum_s \left\{ u_s(\vec{p})b_s(\vec{p})e^{-ip\cdot x} + v_s(\vec{p})d_s(\vec{p})^\dagger e^{ip\cdot x} \right\} , \end{aligned} \tag{6.179}$$

i.e., Eq. (6.173). We now promote $b_s(\vec{p})$ and $d_s(\vec{p})^\dagger$ to operators by imposing the appropriate *anticommutation* relations for the creation and annihilation operators of fermionic particles,

$$\begin{aligned} &\{b_s(\vec{p}), b_t(\vec{q})\} = \{b_s(\vec{p})^\dagger, b_t(\vec{q})^\dagger\} = \{d_s(\vec{p}), d_t(\vec{q})\} = \{d_s(\vec{p})^\dagger, d_t(\vec{q})^\dagger\} = 0 \,, \\ &\{b_s(\vec{p}), b_t(\vec{q})^\dagger\} = \{d_s(\vec{p}), d_t(\vec{q})^\dagger\} = (2\pi)^3 2p^0 \delta^{(3)}(\vec{p} - \vec{q}) \,. \end{aligned} \tag{6.180}$$

These imply the following equal-time anticommutation relations,

$$\begin{aligned}
&\{\psi_\alpha(x),\psi_\beta(y)\}_{\rm ET} = \{\psi_\alpha(x)^\dagger,\psi^\dagger_\beta(y)\}_{\rm ET} = 0\,,\\
&\{\psi_\alpha(x),\psi^\dagger_\beta(y)\}_{\rm ET} = \int d\Omega_p \int d\Omega_q \sum_{s,t} \Big[e^{-i(p\cdot x - q\cdot y)} u_{s\,\alpha}(\vec p) u_{t\,\beta}(\vec q)^\dagger \{b_s(\vec p), b_t(\vec q)^\dagger\}_{\rm ET}\\
&\qquad\qquad + e^{i(p\cdot x - q\cdot y)} v_{s\,\alpha}(\vec p) v_{t\,\beta}(\vec q)^\dagger \{d_s(\vec p)^\dagger, d_t(\vec q)\}_{\rm ET}\Big]\\
&\qquad = \int d\Omega_p \sum_s \Big[e^{-ip\cdot(x-y)} u_{s\,\alpha}(\vec p) u_{s\,\beta}(\vec p)^\dagger + e^{ip\cdot(x-y)} v_{s\,\alpha}(\vec p) v_{s\,\beta}(\vec p)^\dagger\Big]\\
&\qquad = \int d\Omega_p \Big[e^{i\vec p\cdot(\vec x-\vec y)}(\not p + m) + e^{-i\vec p\cdot(\vec x-\vec y)}(\not p - m)\Big]\gamma^0\\
&\qquad = \int d\Omega_p\, p^0 \Big[e^{i\vec p\cdot(\vec x-\vec y)} + e^{-i\vec p\cdot(\vec x-\vec y)}\Big] = \delta^{(3)}(\vec x - \vec y)\,.
\end{aligned} \tag{6.181}$$

This shows us how to modify the canonical quantisation approach for the fermion fields. It is easy to see that the Dirac equation can be obtained via a variational principle from the Lagrangian

$$\mathscr{L} = \bar\psi(i\not\partial - m)\psi\,, \tag{6.182}$$

where $\bar\psi \equiv \psi^\dagger\gamma^0$. Indeed, performing variation with respect to $\bar\psi$ we find

$$\frac{\partial\mathscr{L}}{\partial\bar\psi} = (i\not\partial - m)\psi = \partial_\mu \frac{\partial\mathscr{L}}{\partial(\partial_\mu\bar\psi)} = 0\,. \tag{6.183}$$

Taking the derivative with respect to $\partial_0\psi$ we find the canonical momentum conjugate to ψ,

$$\pi = \frac{\partial\mathscr{L}}{\partial(\partial_0\psi)} = i\psi^\dagger\,. \tag{6.184}$$

Imposing canonical anticommutation relations, we finally find

$$\begin{aligned}
&\{\psi_\alpha(x),\psi_\beta(y)\}_{\rm ET} = 0\,, \qquad \{\pi_\alpha(x),\pi_\beta(y)\}_{\rm ET} = -\{\psi_\alpha(x)^\dagger,\psi_\beta(y)^\dagger\}_{\rm ET} = 0\,,\\
&\{\psi_\alpha(x),\pi_\beta(y)\}_{\rm ET} = \{\psi_\alpha(x), i\psi^\dagger_\beta(y)\}_{\rm ET} = i\delta^{(3)}(\vec x - \vec y\,)\,,
\end{aligned} \tag{6.185}$$

from which the commutation relations for the creation and annihilation operators follows.[45] We can then obtain the Hamiltonian via the usual Legendre transform,

$$\begin{aligned}
H &= \int d^3x \left[\pi(x)\partial_0\psi(x) - \bar\psi(x)\left(i\not\partial - m\right)\psi(x)\right]\\
&= \int d^3x \left[i\bar\psi(x)\gamma^0\partial_0\psi(x) - \bar\psi(x)\left(i\not\partial - m\right)\psi(x)\right]\\
&= \int d^3x\, \bar\psi(x)\left(-i\vec\nabla\cdot\vec\gamma + m\right)\psi(x)\,.
\end{aligned} \tag{6.186}$$

[45] A couple of technical remarks. The Lagrangian in Eq. (6.182) is not Hermitian, but it differs from a Hermitian quantity by a total divergence, which does not affect the equations of motion. The fields ψ and $\bar\psi$ in this case are *not* independent fields, like they were in the scalar case. This is shown for example by the vanishing of the canonical momentum that would be associated to $\psi^\dagger$. Even using the Hermitian form of the Lagrangian, one finds constraints between fields and canonical momenta. The appropriate way to deal with this requires the theory of constrained Hamiltonians, which is way beyond the scope of these notes. Luckily enough, the sloppy approach gives the correct results.

Anticipating the appearance of singularities in the product of fields at the same spacetime point, we define the normal-ordered product for fermion fields in analogy with the scalar case. One writes down the expansion in creation and annihilation operators of the fields in the product, and replaces each string of operators with the analogous string where all creation operators are moved to the left and all annihilation operators are moved on the right, but multiplied by a sign equal to the signature of the permutation: if an odd number of transposition was necessary, then this will be minus, if it was even that this will be plus. This can be summarised in the simple relations

$$\begin{aligned} &: b_s(\vec{p}) b_t(\vec{q})^\dagger : = -b_t(\vec{q})^\dagger b_s(\vec{p}) = - : b_t(\vec{q})^\dagger b_s(\vec{p}) : , \\ &: b_s(\vec{p}) b_t(\vec{q}) : = b_s(\vec{p}) b_t(\vec{q}) = -b_t(\vec{q}) b_s(\vec{p}) = - : b_t(\vec{q}) b_s(\vec{p}) : , \end{aligned} \tag{6.187}$$

which also tell us that now the order in which the operators appear will matter for the overall sign of the normal-ordered product. Defining the positive-frequency and negative-frequency parts of ψ and $\bar{\psi}$,

$$\begin{aligned} \psi_+(x) &= \int d\Omega_p \sum_s u_s(\vec{p}) b_s(\vec{p}) e^{-ip\cdot x} , \qquad \psi_-(x) = \int d\Omega_p \sum_s v_s(\vec{p}) d_s(\vec{p})^\dagger e^{ip\cdot x} , \\ \bar{\psi}_+(x) &= \int d\Omega_p \sum_s \bar{v}_s(\vec{p}) d_s(\vec{p}) e^{-ip\cdot x} , \qquad \bar{\psi}_-(x) = \int d\Omega_p \sum_s \bar{u}_s(\vec{p}) b_s(\vec{p})^\dagger e^{ip\cdot x} , \end{aligned} \tag{6.188}$$

we then have for the product of two fields $\psi_{1,2} = \psi, \bar{\psi}$

$$\begin{aligned} : \psi_1(x)\psi_2(y) : &= \psi_{1+}(x)\psi_{2+}(y) - \psi_{2-}(y)\psi_{1+}(x) + \psi_{1-}(x)\psi_{2+}(y) + \psi_{1-}(x)\psi_{2-}(y) \\ &= -\psi_{2+}(y)\psi_{1+}(x) - \psi_{2-}(y)\psi_{1+}(x) + \psi_{1-}(x)\psi_{2+}(y) - \psi_{2-}(y)\psi_{1-}(x) \\ &= - : \psi_2(y)\psi_1(x) : . \end{aligned} \tag{6.189}$$

Minus signs like the one appearing here are ubiquitous when dealing with fermions. We can also work out the simplest case of Wick's theorem for fermions. In general,

$$\psi_1(x)\psi_2(y) =: \psi_1(x)\psi_2(y) : + \{\psi_{1+}(x), \psi_{2-}(y)\} , \tag{6.190}$$

with the anticommutator being a c-number and thus equal to its vacuum expectation value. The only case in which this is nonzero is when one of the fields is ψ and the other is $\bar{\psi}$, while in all other cases the ordinary product and the normal-ordered product are equal. In the nontrivial case

$$\begin{aligned} \psi(x)\bar{\psi}(y) &=: \psi(x)\bar{\psi}(y) : + \{\psi_+(x), \bar{\psi}_-(y)\} \\ &=: \psi(x)\bar{\psi}(y) : + \langle 0 | \{\psi_+(x), \bar{\psi}_-(y)\} | 0 \rangle \\ &=: \psi(x)\bar{\psi}(y) : + \langle 0 | \psi_+(x)\bar{\psi}_-(y) | 0 \rangle \\ &=: \psi(x)\bar{\psi}(y) : + \langle 0 | \psi(x)\bar{\psi}(y) | 0 \rangle . \end{aligned} \tag{6.191}$$

If we define the time-ordered product of two fermion fields as follows,

$$T\left(\psi_1(x)\psi_2(y)\right) = \theta(x^0 - y^0)\psi_1(x)\psi_2(y) - \theta(y^0 - x^0)\psi_2(y)\psi_1(x) = -T\left(\psi_2(y)\psi_1(x)\right) , \tag{6.192}$$

we then conclude

$$T\left(\psi(x)\bar{\psi}(y)\right) =: \psi(x)\bar{\psi}(y) : + \langle 0 | T\left(\psi(x)\bar{\psi}(y)\right) | 0 \rangle \equiv : \psi(x)\bar{\psi}(y) : + S(x-y) . \tag{6.193}$$

The contraction, or propagator, $S(x-y)$ is now a 4×4 matrix, $S=S_{\alpha\beta}$. Dependence on the difference $x-y$ only follows from the translation-invariance of the theory. The generalisation to n fields is like for scalar fields, with a time-ordered prduct being decomposed in a sum of normal-ordered products and contractions, with the difference that now one must multiply each term by the parity of the permutation required to bring the fields in the order in which they appear. For example,

$$\begin{aligned} T\left(\psi_\alpha(x)\bar\psi_\beta(y)\psi_\gamma(w)\bar\psi_\delta(z)\right) = \;& :\psi_\alpha(x)\bar\psi_\beta(y)\psi_\gamma(w)\bar\psi_\delta(z): + S_{\alpha\beta}(x-y):\psi_\gamma(w)\bar\psi_\delta(z): \\ & - S_{\alpha\delta}(x-z):\psi_\gamma(w)\bar\psi_\beta(y): + :\psi_\alpha(x)\bar\psi_\beta(y): S_{\gamma\delta}(w-z) \\ & + S_{\alpha\beta}(x-y)S_{\gamma\delta}(w-z) - S_{\alpha\delta}(x-z)S_{\gamma\beta}(w-y)\,. \end{aligned} \tag{6.194}$$

6.10.3 Interacting theories with fermions

Interacting theories are built in analogy with the scalar case: one adds an interaction term $\mathscr{L}_I$ to the free Dirac Langrangian, solves the euler-Lagrange equations of motion and imposes canonical commutation relations on bosonic fields and anticommutation relations on fermionic fields. This is usually an impossible task, and one resorts to perturbation theory. Fields in the interaction picture are defined like in the bosonic case, and the S-matrix elements can be computed perturbatively making use of Wick's theorem. We consider here an example to illustrate the procedure.

Consider an interaction term $\mathscr{L}_I = g:\phi\bar\psi\psi:$, where ϕ is a Hermitian scalar field. This type of interaction is known as *Yukawa interaction*. Going over to the interaction picture and looking at the elastic scattering of two fermions, we have to lowest perturbative order

$$\langle f|S-\mathbf{1}|i\rangle = \frac{(ig)^2}{2}\int d^4x\int d^4y\langle f|T\left(:\phi(x)\bar\psi(x)\psi(x)::\phi(y)\bar\psi(y)\psi(y):\right)|i\rangle\,, \tag{6.195}$$

where

$$|i\rangle = |\vec p_1\, s_{1\,z};\vec p_2\, s_{2\,z}\rangle\,,\qquad |f\rangle = |\vec p_1'\, s'_{1\,z};\vec p_2'\, s'_{2\,z}\rangle\,. \tag{6.196}$$

Using Wick's theorem for the bosonic fields, we get

$$\langle f|S-\mathbf{1}|i\rangle = \frac{(ig)^2}{2}\int d^4x\int d^4y\, D(x-y)\langle f|T\left(:\bar\psi(x)\psi(x)::\bar\psi(y)\psi(y):\right)|i\rangle\,, \tag{6.197}$$

and since we have as many particles as fermionic fields, only the no-contraction term in the Wick expansion for fermions will contribute,

$$\langle f|S-\mathbf{1}|i\rangle = \frac{(ig)^2}{2}\int d^4x\int d^4y\, D(x-y)\langle f|:\bar\psi(x)\psi(x)\bar\psi(y)\psi(y):|i\rangle\,. \tag{6.198}$$

The annihilation operator required to deal with particles in the initial state appears in ψ, while the creation operator required to deal with particles in the final state appears in $\bar\psi$, and so (understanding the sum over repeated Greek indices)

$$\begin{aligned} &\langle f|:\bar\psi_\alpha(x)\psi_\alpha(x)\bar\psi_\beta(y)\psi_\beta(y):|i\rangle = \langle f|:\bar\psi_\alpha(x)\bar\psi_\beta(y)\psi_\beta(y)\psi_\alpha(x):|i\rangle \\ &= \langle f|:\bar\psi_\alpha(x)\bar\psi_\beta(y):|0\rangle\langle 0|:\psi_\beta(y)\psi_\alpha(x):|i\rangle \\ &= \Big(\langle\vec p_1'\, s'_{1\,z}|\bar\psi_\alpha(x)|0\rangle\langle\vec p_2'\, s'_{2\,z}|\bar\psi_\beta(y)|0\rangle - \langle\vec p_2'\, s'_{2\,z}|\bar\psi_\alpha(x)|0\rangle\langle\vec p_1'\, s'_{1\,z}|\bar\psi_\beta(y)|0\rangle\Big) \\ &\quad\times\Big(\langle 0|\psi_\alpha(x)|\vec p_1\, s_{1\,z}\rangle\langle 0|\psi_\beta(y)|\vec p_2\, s_{2\,z}\rangle - \langle 0|\psi_\alpha(x)|\vec p_2\, s_{2\,z}\rangle\langle 0|\psi_\beta(y)|\vec p_1\, s_{1\,z}\rangle\Big)\,. \end{aligned} \tag{6.199}$$

The minus signs appear due to the anticommuting nature of the creation/annhilation operators. Indeed, one has

$$\begin{aligned}
&b_t(\vec{q})b_s(\vec{p})|\vec{p}_1\,s_{1\,z};\vec{p}_2\,s_{2\,z}\rangle = b_t(\vec{q})b_s(\vec{p})b_{s_{1\,z}}(\vec{p}_1\,)^\dagger|\vec{p}_2\,s_{2\,z}\rangle \\
&= b_t(\vec{q})\left[\delta_{ss_{1\,z}}(2\pi)^3 2p_1^0\delta^{(3)}(\vec{p}-\vec{p}_1) - b_{s_{1\,z}}(\vec{p}_1\,)^\dagger b_s(\vec{p})\right]|\vec{p}_2\,s_{2\,z}\rangle \\
&= \delta_{ss_{1\,z}}(2\pi)^3 2p_1^0\delta^{(3)}(\vec{p}-\vec{p}_1)\delta_{ts_{2\,z}}(2\pi)^3 2p_2^0\delta^{(3)}(\vec{p}-\vec{p}_2)|0\rangle \\
&\qquad - b_t(\vec{q})b_{s_{1\,z}}(\vec{p}_1\,)^\dagger b_s(\vec{p})b_{s_{2\,z}}(\vec{p}_2\,)^\dagger|0\rangle \\
&= \delta_{ss_{1\,z}}(2\pi)^3 2p_1^0\delta^{(3)}(\vec{p}-\vec{p}_1)\delta_{ts_{2\,z}}(2\pi)^3 2p_2^0\delta^{(3)}(\vec{p}-\vec{p}_2)|0\rangle \\
&\qquad - b_t(\vec{q})b_{s_{1\,z}}(\vec{p}_1\,)^\dagger\delta_{ss_{2\,z}}(2\pi)^3 2p_2^0\delta^{(3)}(\vec{p}-\vec{p}_2)|0\rangle \\
&= \delta_{ss_{1\,z}}(2\pi)^3 2p_1^0\delta^{(3)}(\vec{p}-\vec{p}_1)\delta_{ts_{2\,z}}(2\pi)^3 2p_2^0\delta^{(3)}(\vec{p}-\vec{p}_2)|0\rangle \\
&\qquad - \delta_{ts_{2\,z}}(2\pi)^3 2p_1^0\delta^{(3)}(\vec{p}-\vec{p}_1)\delta_{ss_{2\,z}}(2\pi)^3 2p_2^0\delta^{(3)}(\vec{p}-\vec{p}_2)|0\rangle\,.
\end{aligned} \tag{6.200}$$

The one-particle matrix elements are easy to compute, and read

$$\begin{aligned}
\langle 0|\psi(x)|\vec{p}\,s_z\rangle &= \int d\Omega_q \sum_s e^{-iq\cdot x}u_s(\vec{q})\langle 0|b_s(\vec{q})b_{s_z}(\vec{p})^\dagger|0\rangle \\
&= \int d\Omega_q \sum_s e^{-iq\cdot x}u_s(\vec{q})\langle 0|\{b_s(\vec{q}),b_{s_z}(\vec{p})^\dagger\}|0\rangle \\
&= \int d\Omega_q \sum_s e^{-iq\cdot x}u_s(\vec{q})\delta_{ss_z}(2\pi)^3 2q^0\delta^{(3)}(\vec{p}-\vec{q}) = e^{-ip\cdot x}u_{s_z}(\vec{p})\,, \\
\langle \vec{p}\,s_z|\bar{\psi}(x)|0\rangle &= \int d\Omega_q \sum_s e^{iq\cdot x}\bar{u}_s(\vec{q})\langle 0|b_{s_z}(\vec{p})^\dagger b_s(\vec{q})|0\rangle \\
&= \int d\Omega_q \sum_s e^{iq\cdot x}\bar{u}_s(\vec{q})\langle 0|\{b_{s_z}(\vec{p})^\dagger,b_s(\vec{q})\}|0\rangle \\
&= \int d\Omega_q \sum_s e^{iq\cdot x}\bar{u}_s(\vec{q})\delta_{ss_z}(2\pi)^3 2q^0\delta^{(3)}(\vec{p}-\vec{q}) = e^{ip\cdot x}\bar{u}_{s_z}(\vec{p})\,.
\end{aligned} \tag{6.201}$$

Since we are integrating over x and y, they are equivalent and therefore

$$\begin{aligned}
\langle f|S-\mathbf{1}|i\rangle &= \frac{(ig)^2}{2}\int d^4x\int d^4y\,D(x-y) \\
&\quad\times 2\Big(\langle \vec{p}_1'\,s'_{1\,z}|\bar{\psi}_\alpha(x)|0\rangle\langle \vec{p}_2'\,s'_{2\,z}|\bar{\psi}_\beta(y)|0\rangle\langle 0|\psi_\alpha(x)|\vec{p}_1\,s_{1\,z}\rangle\langle 0|\psi_\beta(y)|\vec{p}_2\,s_{2\,z}\rangle \\
&\qquad - \langle \vec{p}_2'\,s'_{2\,z}|\bar{\psi}_\alpha(x)|0\rangle\langle \vec{p}_1'\,s'_{1\,z}|\bar{\psi}_\beta(y)|0\rangle\langle 0|\psi_\alpha(x)|\vec{p}_1\,s_{1\,z}\rangle\langle 0|\psi_\beta(y)|\vec{p}_2\,s_{2\,z}\rangle\Big) \\
&= (ig)^2\int d^4x\int d^4y\,D(x-y) \\
&\quad\times\Big(e^{i(p_1'-p_1)\cdot x}e^{i(p_2'-p_2)\cdot y}\bar{u}_{s'_{1\,z}}(\vec{p}_1'\,)u_{s_{1\,z}}(\vec{p}_1\,)\bar{u}_{s'_{2\,z}}(\vec{p}_2'\,)u_{s_{2\,z}}(\vec{p}_2\,) \\
&\quad - e^{i(p_2'-p_1)\cdot x}e^{i(p_1'-p_2)\cdot y}\bar{u}_{s'_{2\,z}}(\vec{p}_2'\,)u_{s_{1\,z}}(\vec{p}_1\,)\bar{u}_{s'_{1\,z}}(\vec{p}_1'\,)u_{s_{2\,z}}(\vec{p}_2\,)\Big)\,.
\end{aligned} \tag{6.202}$$

Integrating over the position of the vertices we finally obtain

$$\begin{aligned}\langle f|S-\mathbf{1}|i\rangle &= i(2\pi)^4\delta^{(4)}(p_1'+p_2'-p_1-p_2)\\ &\quad\times ig^2\Big(\tilde{D}(p_1'-p_1)\bar{u}_{s_1'\,z}(\vec{p}_1')u_{s_1\,z}(\vec{p}_1)\bar{u}_{s_2'\,z}(\vec{p}_2')u_{s_2\,z}(\vec{p}_2)\\ &\qquad-\tilde{D}(p_2'-p_1)\bar{u}_{s_2'\,z}(\vec{p}_2')u_{s_1\,z}(\vec{p}_1)\bar{u}_{s_1'\,z}(\vec{p}_1')u_{s_2\,z}(\vec{p}_2)\Big),\end{aligned} \tag{6.203}$$

and thus

$$\begin{aligned}\mathcal{M}_{fi} = -g^2\Bigg[&\frac{1}{t-m^2}\bar{u}_{s_1'\,z}(\vec{p}_1')u_{s_1\,z}(\vec{p}_1)\bar{u}_{s_2'\,z}(\vec{p}_2')u_{s_2\,z}(\vec{p}_2)\\ &-\frac{1}{u-m^2}\bar{u}_{s_2'\,z}(\vec{p}_2')u_{s_1\,z}(\vec{p}_1)\bar{u}_{s_1'\,z}(\vec{p}_1')u_{s_2\,z}(\vec{p}_2)\Bigg].\end{aligned} \tag{6.204}$$

The relative minus sign between the two terms reflects the fermionic nature of the colliding particles.

Like in the bosonic case, the calculation of S-matrix element can be sped up by means of diagrammatic techniques. The procedure is the same outlined above in Section 6.9, which must be supplemented by rules to deal with fermion fields. An important difference is that for the Dirac fields the lines representing the fields that appear in an interaction vertex must be oriented. The convention is to have ψ represented by an arrow entering the vertex, and $\bar{\psi}$ by an arrow exiting the vertex. Correspondingly, once that a field is associated with an incoming or outgoing particle, one will have

- a line entering the diagram for each particle in the initial state;
- a line exiting the diagram for each particle in the final state;
- a line exiting the diagram for each antiparticle in the initial state;
- a line entering the diagram for each antiparticle in the final state.

The different rules for the antiparticles come from the fact that they are annihilated by $\bar{\psi}$ in the initial state and by ψ the final state. When making the contraction between two fermion fields, an internal line appears directed from the vertex where we picked $\bar{\psi}$ towards the vertex where we picked ψ. Both when associating fields with incoming/outgoing particles and when contracting two fields, minus signs should multiply the various contribution reflecting the number of transpositions of fermion fields required by the application of Wick's theorem, and by the application of annihilation/creation fields to remove particles from the initial/final state. Of course, the factors associated with the colliding particles are different for fermions, but the same phase factors appear as in the bosonic case, so that integrating over the positions of the vertices, thus going over to momentum space, one finds momentum-conserving delta-functions at each vertex (momenta are assigned to fermionic external lines as in the bosonic case, and to fermionic internal lines according to the direction of their arrow). In momentum space the rules for external lines are thus to include

- $u_s(\vec{p})$ for a particle in the initial state;
- $\bar{u}_s(\vec{p})$ for a particle in the final state;
- $\bar{v}_s(\vec{p})$ for an antiparticle in the initial state;

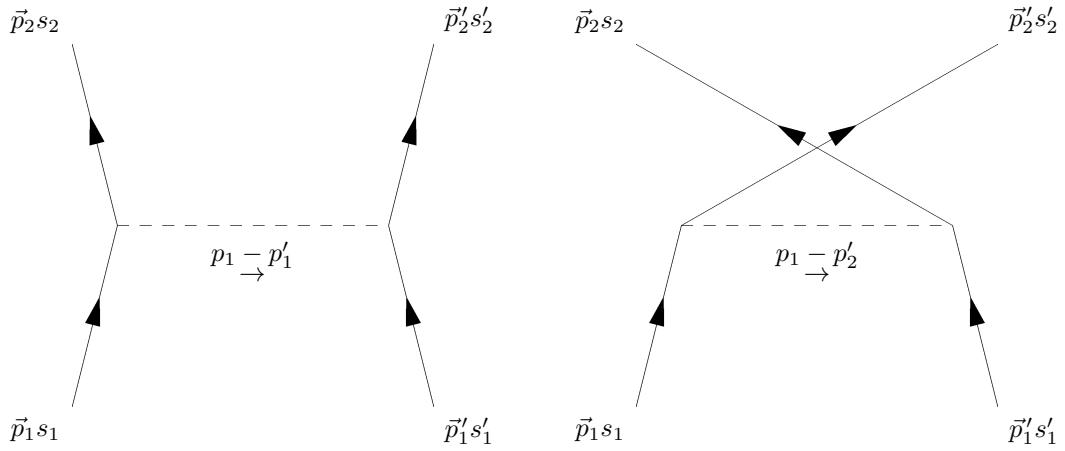

Figure 32: Lowest-order Feynman diagrams for $2 \to 2$ scattering in the Yukawa theory.

- $v_s(\vec{p})$ for an antiparticle in the final state.

The values of $\vec{p}$ and s are chosen according to the state of the colliding particle. To internal lines one associates the Fourier transform $\tilde{S}(q)$ of the fermion propagator. The indices of the bispinors and of the propagators are contracted following the direction opposite to that indicated by the arrows: one starts from the end of an uninterrupted fermion line, corresponding to a final fermion or an initial antifermion, and writes down $\bar{u}_{s\,\alpha_1}(\vec{p})$ or $\bar{v}_{s\,\alpha_1}(\vec{p})$, then keeps placing on the right the fermion propagators $\tilde{S}_{\alpha_i\alpha_{i+1}}(q)$ with q the momentum flowing in the internal line (in the same direction of the arrow), until another initial or final particle is reached, and one closes the line with $u_{s\,\alpha_N}(\vec{p})$ or $v_{s\,\alpha_N}(\vec{p})$ according to whether that is an initial fermion or a final antifermion. With these rules it is straightforward to obtain Eq. (6.204): the two topologically inequivalent Feynman diagrams are shown in Fig. (32). Correspondingly, one writes down a factor $(ig)^2$ for the two vertices, follows the arrows to write the bispinor factors, and multiplies by a scalar propagator. Since the two fermions in the final state are exchanged to go from the first to the second diagram, a relative minus sign appears. Counting is easy: there are two equivalent vertices, which give a factor 2 canceling the 1/2 due to these diagrams being $\mathcal{O}(g^2)$, but no extra degeneracies.

For completeness, we now derive the explicit form of the fermion propagator. From its definition, and making use of the decomposition in creation and annihilation operators, we find

$$\begin{aligned} S_{\alpha\beta}(x-y) &= \langle 0|T\left(\psi_\alpha(x)\bar{\psi}_\beta(y)\right)|0\rangle \\ &= \int d\Omega_p \int d\Omega_q \sum_{s,t} \theta(x^0-y^0) e^{-i(p\cdot x - q\cdot y)} u_{s\,\alpha}(\vec{p})\bar{u}_{t\,\beta}(\vec{q})\langle 0|b_s(\vec{p})b_t(\vec{q})^\dagger|0\rangle \\ &\qquad - \theta(y^0-x^0) e^{-i(p\cdot y - q\cdot x)} v_{t\,\alpha}(\vec{q})\bar{v}_{s\,\beta}(\vec{p})\langle 0|d_s(\vec{p})d_t(\vec{q})^\dagger|0\rangle \\ &= \int d\Omega_p \sum_s \theta(x^0-y^0) e^{-ip\cdot(x-y)} u_{s\,\alpha}(\vec{p})\bar{u}_{s\,\beta}(\vec{p}) \\ &\qquad - \theta(y^0-x^0) e^{-ip\cdot(y-x)} v_{s\,\alpha}(\vec{p})\bar{v}_{s\,\beta}(\vec{p})\,. \end{aligned} \tag{6.205}$$

Using now the sum rules for the bispinors, Eq.(6.168), we get

$$S_{\alpha\beta}(x-y) = \int d\Omega_p \left[\theta(x^0-y^0)e^{-ip\cdot(x-y)}(\not{p}+m)_{\alpha\beta} - \theta(y^0-x^0)e^{-ip\cdot(y-x)}(\not{p}-m)_{\alpha\beta}\right] . \tag{6.206}$$

The momentum-space expression is obtained more easily if we first notice that the fermion propagator obeys the following diferential equation,

$$\begin{aligned}(i\not{\partial}^x - m)S(x-y) &= (i\not{\partial}^x - m)\langle 0|T\left(\psi(x)\bar{\psi}(y)\right)|0\rangle \\ &= \langle 0|T\left((i\vec{\gamma}\cdot\vec{\nabla}_x - m)\psi(x)\bar{\psi}(y)\right)|0\rangle + \partial_0^x\langle 0|T\left(i\gamma^0\psi(x)\bar{\psi}(y)\right)|0\rangle \\ &= \langle 0|T\left((i\not{\partial}^x - m)\psi(x)\bar{\psi}(y)\right)|0\rangle + i\delta(x^0-y^0)\langle 0|\{\gamma^0\psi(x),\bar{\psi}(y)\}_{\rm ET}|0\rangle .\end{aligned} \tag{6.207}$$

But ψ obeys the Dirac equation, and from the canonical anticommutation relations we have

$$\gamma^0_{\alpha\beta}\{\psi_\beta(x),\bar{\psi}_\delta(y)\}_{\rm ET} = \gamma^0_{\alpha\beta}\{\psi_\beta(x),\psi^\dagger_\gamma(y)\}_{\rm ET}\gamma^0_{\gamma\delta} = \delta_{\alpha\delta}\delta^{(3)}(\vec{x}-\vec{y}) , \tag{6.208}$$

and so we conclude that

$$(i\not{\partial}^x - m)S(x-y) = i\delta^{(4)}(x-y) . \tag{6.209}$$

After setting

$$S(x) = \int \frac{d^4q}{(2\pi)^4}\, e^{-iq\cdot x}\tilde{S}(q) , \tag{6.210}$$

we obtain the coresponding equation in momentum space,

$$(\not{q} - m)\tilde{S}(q) = i . \tag{6.211}$$

Inverting the matrix $(\not{q}-m)$ is not possible when it is singular, and since

$$(\not{q}+m)(\not{q}-m) = q^2 - m^2 , \tag{6.212}$$

this is so when $q^2 = m^2$. Away from that point, $(\not{q}-m)^{-1} = (\not{q}+m)/(q^2-m^2)$. The singularity should be dealt with in such a way as to reproduce Eq. (6.206), and we already know that the correct combination of theta functions and plane-wave phases is obtained with

$$\tilde{S}(q) = i\frac{\not{q}+m}{q^2-m^2+i\epsilon} . \tag{6.213}$$

(Be aware that here $q^0 \neq \sqrt{\vec{q}^2+m^2}$, but is a free integration variable). Indeed, use the residue theorem we find

$$\begin{aligned}S(x) &= i\int \frac{d^4q}{(2\pi)^4}\, e^{-iq\cdot x}\frac{\not{q}+m}{q^2-m^2+i\epsilon} \\ &= i\int \frac{d^4q}{(2\pi)^4}\left\{\theta(x^0)\frac{(-2\pi i)}{2q^0}e^{-i(q^0x^0+\vec{q}\cdot\vec{x})}(q^0\gamma^0 - \vec{q}\cdot\vec{\gamma}+m)\right. \\ &\qquad\left. + \theta(-x^0)\frac{2\pi i}{(-2q^0)}e^{-i(-q^0x^0+\vec{q}\cdot\vec{x})}(-q^0\gamma^0 - \vec{q}\cdot\vec{\gamma}+m)\right\}_{q^0=\sqrt{\vec{q}^2+m^2}} \\ &= \int d\Omega_q\left\{\theta(x^0)e^{-iq\cdot x}(\not{q}+m) - \theta(-x^0)e^{iq\cdot x}(\not{q}-m)\right\},\end{aligned} \tag{6.214}$$

in agreement with Eq. (6.206).

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 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 spedd of light $c = 1$. Usually, Greek indices run over $0, \ldots, 4$ and Latin indeces 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* $X_\mu = X^\nu g_{\nu\mu}$, 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 $g^{\mu\rho} g_{\rho\nu} = \delta^\mu{}_\nu$, which in the case at hand is $g^{\mu\nu} = g_{\mu\nu}$. 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 $\Delta s^2 < 0$, we speak of a *spacelike* interval; for $\Delta s^2 = 0$, we speak of a *lightlike* or *null* interval. In general, for $X^2 > 0$ we speak of a timelike vector, for $X^2 < 0$ we speak of a spacelike vector, and for $X^2 = 0$ we speak of a timelike 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 the *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

$$\begin{aligned}(X_1 + X_2)^2 &= (X_1)^2 + (X_2)^2 + 2X_1 \cdot X_2 (X_1)^2 + (X_2)^2 + 2(X_1^0 X_2^0 - \vec{X}_1 \cdot \vec{X}_2) \\ &> 2(X_1^0 X_2^0 - \vec{X}_1 \cdot \vec{X}_2) \geq 2(X_1^0 X_2^0 - |\vec{X}_1||\vec{X}_2|)\,,\end{aligned} \tag{A.7}$$

where we made use of the Schwartz inequality. Next, since $X_{1,2}$ are inside the forward lightcone, $(X)^2_{1,2} > 0$ and $X^0_{1,2} > 0$, so $X^0_{1,2} > |\vec{X}_{1,2}|$, and therefore $X_1^0 X_2^0 - |\vec{X}_1||\vec{X}_2| \geq 0$.

$$(X_1 + X_2)^2 > 2(X_1^0 X_2^0 - X_1^0 X_2^0) \geq 0\,, \tag{A.8}$$

i.e., $(X_1 + X_2)^2 > 0$, and obviously $X_1^0 + X_2^0 > 0$.

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 the interval invariant:

$$(X' - Y')^2 = (X - Y)^2 \Rightarrow X'^2 + Y'^2 - 2X' \cdot Y' = X^2 + Y^2 - 2X \cdot Y \Rightarrow X' \cdot Y' = X \cdot Y\,. \tag{A.9}$$

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*. 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. 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

$$\mathbf{\Lambda}^{-1}_{\alpha\beta} = g^{\alpha\mu} \Lambda^\nu{}_\mu g_{\nu\beta} = \Lambda_\beta{}^\alpha\,. \tag{A.14}$$

Consider now the $\alpha = 0$, $\beta = 0$ component of Eq. (A.11). We have

$$1 = \Lambda^0{}_0 \Lambda^0{}_0 - \Lambda^i{}_0 \Lambda^i{}_0\,, \tag{A.15}$$

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*. A *proper*

orthocronous Lorentz transformation does not change the direction of time or the orientation of space. 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

$$g^{\alpha\beta} = g^{\mu\nu}\Lambda_\mu{}^\alpha \Lambda_\nu{}^\beta \,. \tag{A.16}$$

Using this relation for the inverse transformation Λ^{-1},

$$g^{\alpha\beta} = g^{\mu\nu}\Lambda_\mu{}^\alpha \Lambda_\nu{}^\beta \,, \tag{A.17}$$

a relation analogous to Eq. (A.15) follows,

$$1 = \Lambda^0{}_0\Lambda^0{}_0 - \Lambda^i{}_i\Lambda^i{}_i \,. \tag{A.18}$$

We now have for X'^0

$$X'^0 = \Lambda^0{}_0 X^0 + \Lambda^0{}_i X^i \,. \tag{A.19}$$

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.18) $|\vec{\Lambda}| < \Lambda^0{}_0$, so that

$$X'^0 > \Lambda^0{}_0 X^0 - \Lambda^0{}_0 X^0 = 0 \,. \tag{A.20}$$

The proper orthocronous Lorentz transformations consists of three-dimensional rotations and boosts. they can be obtained from the rotations and the boosts in, say, the x direction,

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.21}$$

where $\beta = \frac{v}{c} = v$ (we work in natural units where $c = 1$), and $\gamma = 1/\sqrt{1-\beta^2}$. All the other Lorentz transformations are obtained from the proper orthocronous one by means of a *parity* transformation P or a *time reversal* T,

$$P^\mu{}_\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} , \qquad T^\mu{}_\nu = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{A.22}$$

A.2 Point particles

Consider a point particle (with positive mass $m > 0$), travelling in space. Its trajectory is described by the four-vector

$$X^\mu(t) = (ct, \vec{x}(t)) = (t, \vec{x}(t)) \,, \tag{A.23}$$

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,

$$dX^\mu(t) = (dt, d\vec{x}(t)) = dt(1, \tfrac{d\vec{x}}{dt}(t)) = dt(1, \vec{v}(t)) \,, \tag{A.24}$$

where $\vec{v}$ is the particle velocity. The infinitesimal interval $(dX)^2$ is

$$(dX)^2 = dX^\mu dX_\mu = dt^2(1-\vec{v}^2) > 0 \tag{A.25}$$

since $\vec{v}^2 < 1$, i.e., the speed is bounded by the speed of light ($c = 1$). To see this, notice that in the rest frame of the particle, the spacetime trajectory reads trivially

$$X^\mu_R = (\tau, \vec{0}\,,) \tag{A.26}$$

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

$$(dX_R)^2 = d\tau^2 = (dX)^2 = dt^2(1-\vec{v}^2)\,. \tag{A.27}$$

This shows that $\vec{v}^2$ must be smaller than 1, since $d\tau^2$ is positive. Setting $\gamma = \frac{1}{1-\vec{v}^2}$, we find that

$$d\tau^2 = \frac{dt^2}{\gamma^2}\,. \tag{A.28}$$

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 $\vec{v}^2$:

$$\tau = \int d\tau = \int_{t_0}^{t} dt'\,\sqrt{1-\vec{v}^2(t')}\,. \tag{A.29}$$

This is the origin of the twins' paradox.

Under orthocronous Lorentz transformations, τ is an invariant, while X^μ, and smiilarly 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. We then define the four-velocity

$$u^\mu \equiv \frac{dX^\mu}{d\tau} = (\tfrac{dt}{d\tau}, \tfrac{d\vec{x}}{d\tau}) = (\gamma, \gamma\tfrac{d\vec{x}}{dt}) = (\gamma, \gamma\vec{v}) = (\gamma, \gamma\vec{\beta})\,, \tag{A.30}$$

where $\vec{\beta} = \vec{v}/c = \vec{v}$ (in natural units). The four-momentum p^μ of the particle is obtained by multiplying the four-velocity by its mass, which is also an invariant, so that p^μ is again a four-vector,

$$p^\mu \equiv m u^\mu = (\gamma m, \gamma m\vec{\beta})\,. \tag{A.31}$$

Component-wise,

$$\begin{aligned} p^0 &= m\gamma = \frac{m}{\sqrt{1-\vec{v}^2}} = E\,, \\ p^i &= m\gamma\vec{\beta}^{\,i} = \frac{m\vec{v}^{\,i}}{\sqrt{1-\vec{v}^2}} = \vec{p}^{\,i}\,, \end{aligned} \tag{A.32}$$

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

$$\begin{aligned} p^0 &= mc\frac{1}{\sqrt{1-\left(\frac{\vec{v}}{c}\right)^2}} = mc\left(1+\frac{1}{2}\left(\frac{\vec{v}}{c}\right)^2\right), \\ \vec{p}^{\,i} &= mc\frac{\vec{v}}{c}\frac{1}{\sqrt{1-\left(\frac{\vec{v}}{c}\right)^2}} = m\vec{v}+\mathcal{O}((v/c)^2)\,. \end{aligned} \tag{A.33}$$

The second line is, to leading order, the familiar expression for the spatial moemntum of a particle. Multiplying the first line by c, we obtain

$$p^0 c = mc^2 + \frac{1}{2}\vec{v}^2\,, \tag{A.34}$$

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,

$$p^\mu = m\frac{dX^\mu}{d\tau} = (\tfrac{E}{c}, \vec{p})\,. \tag{A.35}$$

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:

$$p^2 = m^2\gamma^2(1-\vec{\beta}^{\,2}) = m^2\,. \tag{A.36}$$

The four-velocity square is simply

$$u^2 = \gamma^2(1-\vec{\beta}^{\,2}) = 1\,. \tag{A.37}$$

Both are timelike vectors; since $u^0 \geq 1$. both u^μ and p^μ are inside the forwards 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,

$$E^2 = \vec{p}^{\,2} + m^2\,. \tag{A.38}$$

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.,

$$\omega^2 - \vec{k}^{\,2} = 0 \Rightarrow \omega = |\vec{k}| \geq 0\,. \tag{A.39}$$

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 particle, c and d (possibly equal to a and b), in the final state,

$$a\,b \to c\,d\,. \tag{A.40}$$

Let us discuss the kinematics of the final state both in the lab frame, in which particle b is at rest, and in the center of mass frame, in which the toal spatial momentum vanishes.

Lab frame In the lab frame the four-momenta of a and b read

$$p_a = (E_L, \vec{p}_L)\,, \qquad p_b = (m_b, 0)\,. \tag{A.41}$$

From now on we drop the coordinate index from the four-vectors. For the particles in the final state we have in general

$$p_c = (E_c, \vec{p}_c)\,, \qquad p_d = (E_d, \vec{p}_d)\,. \tag{A.42}$$

We define the angle θ_L as the angle between the trajectory of c and that of a, from the equation

$$\cos\theta_L = \frac{\vec{p}_L \cdot \vec{p}_c}{|\vec{p}_L||\vec{p}_c|}\,. \tag{A.43}$$

CM frame The CM frame is by definition the frame in which the total spatial moemntum vanishes. Therefore, the four-momenta of the various particles read

$$\begin{aligned} p_a &= (E_a^*, \vec{p}^{\,*})\,, \qquad p_b = (E_b^*, -\vec{p}^{\,*})\,, \\ p_c &= (E_c^*, \vec{p}^{\,\prime *})\,, \qquad p_d = (E_d^*, -\vec{p}^{\,\prime *})\,. \end{aligned} \tag{A.44}$$

Also in this case we define the angle θ^* as the one formed by the trajectories of a and c,

$$\cos\theta^* = \frac{\vec{p}^{\,*} \cdot \vec{p}^{\,\prime *}}{|\vec{p}^{\,*}||\vec{p}^{\,\prime *}|}\,. \tag{A.45}$$

We also denote the total center of mass energy as

$$\sqrt{s} = E_a^* + E_b^* = E_c^* + E_d^*\,. \tag{A.46}$$

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:

$$\begin{aligned} p_a + p_b &= p_c + p_d \\ p_b &= p_c + p_d - p_a \\ p_b^2 &= (p_c + p_d)^2 + p_a^2 - 2p_a \cdot (p_c + p_d) \\ m_b^2 &= s + m_a^2 - 2E_a^*\sqrt{s} \\ E_a^* &= \frac{s + m_a^2 - m_b^2}{2\sqrt{s}}\,. \end{aligned} \tag{A.47}$$

Notice that since s is a relativistic invariant, through this relation we can determine E_a^* from knowledge of E_L in the lab:

$$s = (p_a + p_b)^2 = m_a^2 + m_b^2 + 2p_a \cdot p_b = m_a^2 + m_b^2 + 2E_L m_b\,. \tag{A.48}$$

On the other hand, a derivation entirely analogous to that in Eq. (A.47) with a and b exchanged allows to derive E_b^*,

$$E_b^* = \frac{s + m_b^2 - m_a^2}{2\sqrt{s}}\,, \tag{A.49}$$

and more importantly exchanging a with c and b with d we obtain the energies of the final products,

$$E_c^* = \frac{s + m_c^2 - m_d^2}{2\sqrt{s}}\,, \qquad E_d^* = \frac{s + m_d^2 - m_c^2}{2\sqrt{s}}\,. \tag{A.50}$$

From the dispertion relation we can then derive the magnitude of the momenta:

$$\begin{aligned} |\vec{p}^*|^2 &= E_a^{*2} - m_a^2 = \frac{(s + m_a^2 - m_b^2)^2 - 4sm_a^2}{4s} = \frac{s^2 + (m_a^2 - m_b^2)^2 - 2s(m_a^2 + m_b^2)}{4s} \\ &= \frac{(s + m_a^2 + m_b^2)^2 - 4m_a^2 m_b^2}{4s}\,, \\ |\vec{p}'^*|^2 &= E_c^{*2} - m_c^2 = \frac{(s + m_c^2 - m_d^2)^2 - 4sm_c^2}{4s} = \frac{s^2 + (m_c^2 - m_d^2)^2 - 2s(m_c^2 + m_d^2)}{4s} \\ &= \frac{(s + m_c^2 + m_d^2)^2 - 4m_c^2 m_d^2}{4s}\,. \end{aligned} \tag{A.51}$$

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

$$0 = \gamma_{\rm CM}(|\vec{p}_{\rm lab}| - \beta_{\rm CM} E_{\rm lab})\,, \tag{A.52}$$

where $\vec{p}_{\rm lab}$ and $E_{\rm lab}$ are the total spatial momentum ant total energy in the lab, and so that velocity of the CM in the lab is

$$\beta_{\rm CM} = \frac{|\vec{p}_L|}{E_L + m_b}\,. \tag{A.53}$$

Writing the inverse Lorentz transformation from the CM to the lab we then find

$$\begin{aligned} E_{c,\rm lab} &= \gamma_{\rm CM}(E_c^* + \beta_{\rm CM}|\vec{p}'^*|\cos\theta^*)\,, \\ |\vec{p}_{c,\rm lab}|\cos\theta_L &= \gamma_{\rm CM}(|\vec{p}'^*|\cos\theta^* + \beta_{\rm CM} E_c^*)\,, \\ |\vec{p}_{c,\rm lab}|\sin\theta_L &= |\vec{p}'^*|\sin\theta^*\,, \end{aligned} \tag{A.54}$$

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

$$s = (p_p + p_{\bar{p}})^2 = 2(m_p^2 + E_L m_p) \tag{A.55}$$

where E_L is the energy of the $\bar{p}$ in the lab. Solving for E_L and imposing $\sqrt{s} = 540$ GeV,

$$E_L = \frac{s - 2m_p^2}{2m_p} \simeq \frac{s}{2m_p} \simeq \frac{(540)^2}{2}\ \text{GeV} \simeq \frac{30}{2}\cdot 10^4\ \text{GeV} = 150\ \text{TeV}\,, \tag{A.56}$$

which is a huge energy. In general, the lab energy scales like $E_L \simeq E_{\rm CM}^2/m_p$.

A.4 Mandelstam variables

A convenient set of variables to describe the kinematics of $2 \to 2$ scattering process are the so-called *Mandelstam variables*,

$$\begin{aligned} s &\equiv (p_a + p_b)^2 = (p_c + p_d)^2 \,, \\ t &\equiv (p_a - p_c)^2 = (p_b - p_d)^2 \,, \\ u &\equiv (p_a - p_d)^2 = (p_b - p_c)^2 \,. \end{aligned} \tag{A.57}$$

The main advantage of these variables is that they are Lorentz-invariant by construction. We have already seen that s is the total center of mass energy squared. The variable t is instead the square of the four-momentum transfer from a to c, and reads explicitly

$$t = p_a^2 + p_c^2 - 2p_a \cdot p_c = m_a^2 + m_c^2 - 2(E_a^* E_c^* - |\vec{p}^*||\vec{p}'^*| \cos\theta^*) \,. \tag{A.58}$$

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.58). 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

$$\begin{aligned} s + t + u &= (p_a + p_b)^2 + (p_a - p_c)^2 + (p_a - p_d)^2 \\ &= m_a^2 + m_b^2 + m_c^2 + m_d^2 + 2p_a \cdot (p_a + p_b - p_c - p_d) = m_a^2 + m_b^2 + m_c^2 + m_d^2 \,. \end{aligned} \tag{A.59}$$

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 easy to see that

$$s \geq \max((m_a + m_b)^2, (m_c + m_d)^2) \,, \tag{A.60}$$

while bounds on t and u are in general more complicated to derive. There is however a big simplification 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

$$t = m_a^2 + m_c^2 - \frac{s}{2}\left(1 - \cos\theta^* \sqrt{1 - \frac{4m_a^2}{s}} \sqrt{1 - \frac{4m_c^2}{s}}\right) . \tag{A.61}$$

Moreover, if one also has $m_a = m_c \equiv m$, then

$$t = 2m^2 - \frac{s}{2}\left(1 - \cos\theta^* \left(1 - \frac{4m^2}{s}\right)\right) = -\left(s - 4m^2\right) \sin^2 \frac{\theta^*}{2} \,, \tag{A.62}$$

which, since $s \geq 4m^2$, shows that

$$-\left(s - 4m^2\right) \leq t \leq 0 \,. \tag{A.63}$$

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 that one finds 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

$$\sin^2\frac{\theta^*}{2} = -\frac{t}{s-4m_p^2} = \frac{1.34}{53^2 - 4*0.938^2} = \frac{1.34}{2805} = 4.78\cdot 10^{-4}\,. \tag{A.64}$$

To leading order, $\sin^2\frac{\theta^*}{2} \simeq \frac{(\theta^*)^2}{4}$, and

$$\theta^* \simeq 2\sqrt{5}\cdot 10^{-2} \simeq 4\cdot 10^{-2}\,. \tag{A.65}$$

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 $\mathbb{R}^4$,

$$\Phi = \{p \in \mathbb{R}^4 | p^2 - m^2 = 0\,,\quad p^0 > 0\}\,, \tag{A.66}$$

is called the *phase space* of the particle. The infinitesimal element of phase space has measure

$$d\Phi = d^4p\,\delta(p^2 - m^2)\theta(p^0)C\,, \tag{A.67}$$

where C is an arbitrary constant, usually set to $C = (2\pi)^{-3}$. What matters here is that $d\Phi$ is invariant under orthocronous Lorentz transformations.[46] In fact, p^2 is invariant under generic Lonretz transformations, while the sign of hte 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

$$\delta(f(x)) = \sum_{x_n, f(x_n)=0} \frac{1}{|f'(x_n)|}\delta(x - x_n)\,, \tag{A.68}$$

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

$$\int_{-\infty}^{+\infty} dx\,\delta(f(x))h(x) = \sum_k \int_{I_k} dx\,\delta(f(x))h(x) = \sum_k \int_{f(I_k)} dy\,\frac{1}{|f'(f^{-1}(y))|}\delta(y)h(f^{-1}(y))\,. \tag{A.69}$$

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.69) that contain a zero x_n, and

$$\int_{-\infty}^{+\infty} dx\,\delta(f(x))h(x) = \sum_k \int_{f(I_k)} dy\,\frac{1}{|f'(f^{-1}(0))|}\delta(y)h(f^{-1}(0)) = \sum_n \frac{1}{|f'(x_n)|}h(x_n)\,. \tag{A.70}$$

[46] Since Φ is a noncompact domain, the integral of $d\Phi$ will diverge anyway.

This is precisely what one obtains straightforwardly by repeating the procedure with the right-hand side.

Using the result above, we find

$$\begin{aligned} d\Phi &= d^4p\delta(p^2-m^2)\theta(p^0)C = Cd^4p\delta(p^{0\,2}-\vec{p}^2-m^2) \\ &= Cd^4p\frac{1}{2|p^0|}\left[\delta(p^0-\varepsilon(\vec{p}))+\delta(p^0+\varepsilon(\vec{p}))\right]\theta(p^0) = Cd^4p\frac{1}{2\varepsilon(\vec{p})}\delta(p^0-\varepsilon(\vec{p}))\theta(p^0)\,, \end{aligned} \tag{A.71}$$

where

$$\varepsilon(\vec{p}) \equiv \sqrt{\vec{p}^2+m^2}\,. \tag{A.72}$$

The n-particle phase space is the phase space of n particles subjected to a constraint on the total four-momentum:

$$d\Phi^{(n)} = \prod_{j=1}^{n} d\Phi_j \delta^{(4)}\left(p_{\rm tot}-\sum_{j=1}^{n}p_j\right)\,. \tag{A.73}$$

Let us work out explicitly the two-particle phase space element, which up to a constant reads

$$\begin{aligned} d\Phi^{(2)} &\propto \frac{d^3p_1}{2\varepsilon_1(\vec{p}_1\,)}\frac{d^3p_2}{2\varepsilon_2(\vec{p}_2\,)}\delta^{(4)}(p_{\rm tot}-p_1-p_2) \\ &= \frac{d^3p_1}{2\varepsilon_1(\vec{p}_1\,)}\frac{d^3p_2}{2\varepsilon_2(\vec{p}_2\,)}\delta^{(3)}(\vec{p}_{\rm tot}-\vec{p}_1-\vec{p}_2)\delta(E_{\rm tot}-\varepsilon_1(\vec{p}_1\,)-\varepsilon_2(\vec{p}_2\,))\,, \end{aligned} \tag{A.74}$$

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

$$d\Phi^{(2)} \propto \frac{d^3p_1}{2\varepsilon_1(\vec{p}_1\,)}\frac{1}{2\varepsilon_2(\vec{p}_{\rm tot}-\vec{p}_1\,)}\delta(E_{\rm tot}-\varepsilon_1(\vec{p}_1\,)-\varepsilon_2(\vec{p}_{\rm tot}-\vec{p}_1\,))\,. \tag{A.75}$$

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 center 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)

$$E_{\rm tot}-\varepsilon_1(p)-\varepsilon_2(p)\,, \tag{A.76}$$

and

$$\left|\frac{\partial}{\partial p}\left[E_{\rm tot}-\varepsilon_1(p)-\varepsilon_2(p)\right]\right| = \left[\frac{p}{\varepsilon_1(p)}+\frac{p}{\varepsilon_2(p)}\right]\,. \tag{A.77}$$

Using this in Eq. (A.75), changing variables to $d^3p_1 = dpp^2d\Omega_{\rm CM} = dpp^2d\cos\theta_{\rm CM}d\phi_{\rm CM}$, and integrating over p we find

$$\begin{aligned} d\Phi^{(2)} &\propto \frac{dpp^2d\Omega}{2\varepsilon_1(p)}\frac{1}{2\varepsilon_2(p)}\left[\frac{p}{\varepsilon_1(p)}+\frac{p}{\varepsilon_2(p)}\right]^{-1}\delta(p-p_{\rm CM}) \\ &= d\Omega_{\rm CM}\frac{p_{\rm CM}}{4(\varepsilon_1(p_{\rm CM})+\varepsilon_2(p_{\rm CM}))} = d\Omega_{\rm CM}\frac{p_{\rm CM}}{4E_{\rm CM,tot}} = d\Omega_{\rm CM}\frac{p_{\rm CM}}{4\sqrt{s}}\,. \end{aligned} \tag{A.78}$$

References

[1] David J. Griffiths, *Introduction to elementary particles*, 2nd edition (Wiley, New York, 2008).

[2] Steven Weinberg, *The quantum theory of fields. Vol. 1: Foundations* (Cambridge University Press, Cambridge, 1995).

[3] Morton Hamermesh, *Group theory and its application to physical problems*, (Addinson-Wesley, Reading, 1962).