Introduction to the Standard Model of Particle Physics – Part 2

Rainer F. Hauser
rainer.hauser@gmail.com

August 22, 2019

Abstract

Despite its limits, the so-called Standard Model describes large parts of Particle Physics in very exact accordance with experiments and is very well tested. Alex Flournoy from the Colorado School of Mines held 29 lectures in 2018 covering this topic. His lectures were available on YouTube at the time this transcript has been assembled and may still be available today as "Particle Physics (2018)" in 29 separate videos. The first 16 lectures cover the theoretical side of the topic with the formal structure of the Standard Model, symmetries, gauge fields and the Higgs mechanism, and the remaining 13 lectures touch on the practical side with the computational aspects of decays and scattering amplitudes. A thorough basis of modern physics including Special Relativity and Quantum Mechanics is required, but knowledge of Quantum Field Theory is not a prerequisite.

1 Introduction

1.1 Short Review of the Standard Model

To learn today's view of Particle Physics through a historical approach is very difficult. Especially during the period from the fifties to the eighties of last century there was a lot of confusion and uncertainty. It is much easier to understand what is going on in the Standard Model first and only then go back to review its history.

Symmetries play an important role in the Standard Model. There is the group SO(1,3) of spacetime and the invariance under the group of Lorentz transformations in Special Relativity which is important here because massless particles need it always and massive particles need it when accelerated for scattering experiments to a speed close to the speed of light. In addition the groups SU(3), SU(2), U(1) play an important role for the strong, the weak and the electromagnetic force. Every single particle has got to fall into some representation of all these four groups. The electron, for example, is a spinor in spacetime, a singlet under the strong interaction, part of a doublet together with the electron-neutrino under the weak interaction, and it is charged under electromagnetism. The quark, as another example, is also a spinor in spacetime, part of triplets under the strong interaction, part of a doublet under the weak interaction and they also carry electric charge.

It turned out that all the gauge fields are spin-1 fields and fulfill the Proca equation, all the matter particles are spin-$\frac{1}{2}$ fields and fulfill the Dirac equation, and the Higgs field is a spin-0 field and fulfills the Klein-Gordon equation. In contrast to experience, all gauge bosons appearing from gauge fields when localizing a symmetry must be massless, and because of the fact that the weak interactions treat fermions with left chirality differently than those with right chirality also the fermions must all be massless. The Higgs mechanism solves both these problems with the missing mass.

The Standard Model ignores gravity because the strong and weak interactions do not matter where gravity matters, and gravity does not matter where the strong and weak interactions matter. In the course of a Particle Physics experiment or in the course of an electron orbiting a nucleus, one has never to take the gravitational effects into account. The masses are too tiny. Only for black holes all four forces are important, and that is one of the reasons why black holes are studied so intensively. Also in Cosmology one has the use General Relativity because the universe is certainly curved, but also Particle Physics play an important role because vacuum energies get created and give contributions to the energy-momentum tensor which curves the space.

The theory of the Standard Model so far gave an understanding why the Standard Model Lagrangian looks the way it does. It got a Klein-Gordon part because of the Higgs boson, it got some Dirac parts because all the matter particles are spinors, it got some Proca parts for all the gauge fields, and it finally got terms for interactions.

1.2 Conventions

So far the theory was more based on a Quantum Field Theory, but calculations are more associated with Particle Physics. Resources in Particle Physics use a different set of conventions than Quantum Field Theory. Thus, here these conventions are adopted.

The metric in Minkowski space is

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

with the consequence that the dual vector is $V_\mu = (V^0, -V^1, -V^2, -V^3)$ for a vector V^μ with components V^0, V^1, V^2, V^3 and therefore also that $V^\mu V_\mu = (V^0)^2 - (V^1)^2 - (V^2)^2 - (V^3)^2$.

Together with the definitions $\bar{\Psi} = \Psi^\dagger \gamma^0$, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and the mass-shell condition $E^2 - \left|\vec{P}\right|^2 c^2 = m^2 c^4$ also the Lagrangians and the equations of motion for free fields change to

$$\begin{aligned} \mathcal{L}_{\text{Klein-Gordon}} &= \frac{1}{2}\partial_\mu \Phi \partial^\mu \Phi - \frac{1}{2}\left(\frac{mc}{\hbar}\right)^2 \Phi^2 & \partial_\mu \partial^\mu \Phi + \left(\frac{mc}{\hbar}\right)^2 \Phi &= 0 \\ \mathcal{L}_{\text{Dirac}} &= \hbar c \bar{\Psi} \gamma^\mu \partial_\mu \Psi - mc^2 \bar{\Psi}\Psi & i\gamma^\mu \partial_\mu \Psi - \frac{mc}{\hbar}\Psi &= 0 \\ \mathcal{L}_{\text{Proca}} &= \frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} + \frac{1}{8\pi}\left(\frac{mc}{\hbar}\right)^2 A_\mu A^\mu & \partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) + \left(\frac{mc}{\hbar}\right)^2 A^\nu &= 0 \end{aligned}$$

because of the use of the different Minkowski metric.

2 Basics of Decays and Collisions

2.1 Some Definitions

In Particle Physics experiments there are decays where a massive particle disappears and lighter particles come out and there is scattering where one takes more than one particle and shoot them to or near each other such that they interact and some particles come out. Decays change the flavor of the particles, and the only force which can do that is the weak interaction. Particles are tiny fluctuations of the fields. When a particle decays, it can often decay in many different possibilities. A particle A can decay into two particles B and C, into other two particles D and E, into three particles F, G and H and so on. The decay of A into F, G and H is written as $A \to F + G + H$ and similarly for the other examples. The total mass on the right side of the arrow cannot be bigger than the mass on the left side of the arrow because one cannot create energy by a decay. Therefore the lightest particle cannot decay and is stable.

The list of possible decays does not include $A \to B$ although a term with some constants times $\eta\beta$ belong to processes like where a η all of a sudden turns into a β, but this is not an interaction in the canonical sense. It just means that η and β are not distinguishable. Decays lead always to two or more particles.

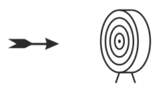
Each of the possibilities $A \to B + C$, $A \to D + E$ and so on is called a *decay channel*. For each channel will be a *decay rate* Γ_i where the decay rate is the probability per unit time of A decaying into this possibility. The total decay rate Γ_{tot} is the sum over all individual decay rates Γ_i and is the probability per unit time of A decaying at all. If one has a sample of the same particles and wants to know the number dN of that sample changing with time, then

$$dN = -\Gamma_{\text{tot}}\, N\, dt \qquad\qquad N(t) = N(0)\, e^{-\Gamma_{\text{tot}} t} \qquad\qquad \tau_{\text{avg}} = \frac{1}{\Gamma_{\text{tot}}} \tag{2.1}$$

where τ_{avg} is the average lifetime. The minus sign in the left equation indicates that the number N decreases with time. The different Γ_i can be calculated, and if they are all known also the other quantities Γ_{tot}, τ_{avg} and $N(t)$ can be determined.

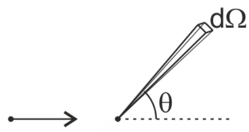
For collisions, the likelihood of a particular collision event $A + B \to C + D$ where $C + D$ is the outcome is the *scattering cross-section* σ_i. The total or inclusive cross-section for $A + B$ is the sum σ_{tot} over the cross-sections σ_i of all possible outcomes for $A + B$ which are called *scattering channels*. The likelihood that $A + B$ interact at all is σ_{tot}.

A scattering like firing an arrow at a target has a likelihood which is determined by the actual cross-sectional area of the target. In Particle Physics scattering is, however, much more complicated with many different types of collisions:



- soft target (interaction with potential)
- depends on identity of arrow (electron or neutrino, for example)
- velocity dependent
- multiple ways to successfully hit
- arrow: final state "hit" or "no hit"; particles: many possible outcomes

When one actually does experiments, often the scattering cross-section is not easy to detect. If the beam hits the target, usually particles come out in all directions. Detectors in some places are more effective than in others, and sometimes the view is limited to a small slice of solid angle $d\Omega$ where the detector is located. Thus, one might instead need $\frac{d\sigma}{d\Omega}$ as the differential scattering cross-section which typically only depends on θ and not on the angle about the direction of the beam. If one knows $\frac{d\sigma}{d\Omega}$ one can integrate over all $d\Omega$ and gets back the total σ.



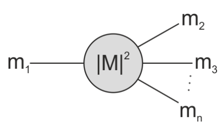
The interest here is in relativistic and quantum mechanical calculations of the decay rates Γ_i and scattering cross-sections σ_i. This would really entail full Quantum Field Theory, but one can skip some steps and try to make sense of the results without a complete derivation.

2.2 Fermi's Golden Rule

For decays and scattering the Γ_i and the σ_i depend roughly on two factors. The first factor is kinematic considerations with the phase space degrees of freedom including conservation of energy and restriction to positive energy. If, for example, a very massive object decays into another object which is very close in mass then the amount of phase space degrees of freedom is smaller than if this object decays into something very light because a massive object decaying into something very light liberates a lot of extra energy which can be distributed in many different ways leading to a lot of phase space to work with. This is reflected in the likelihood. The second factor is dynamics with the particular interactions (strong interactions, weak interactions) and intermediate states (virtual particles). The intermediate states do not have to satisfy the same kinematic constraints as the real physical observable states.

The kinematic contribution to Γ_i and σ_i is summed up in Fermi's golden rule which works for any interaction. There is a version of it for decays and a version for scattering. There is one quantity within Fermi's golden rule that is still undetermined which turns out to be the quantum amplitude for the process. To calculate it the second factor dynamics comes into the game. Here the Feynman calculus with the Feynman diagrams plays a crucial role.

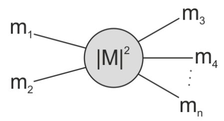
For the decay version of the golden rule it is assumed that a particle with mass m_1 is at rest and decays into $n - 1$ particles with mass m_2, ... , m_n in channel i. Because the initial mass m_1 and the resulting masses m_2, ... , m_n together do not have to be the same, some energy of mass will be liberated and turned into kinetic energy such that not all resulting particles will be at rest.



The decay process for channel i is therefore $m_1^{\text{rest}} \to m_2 + m_3 + ... + m_n$. The corresponding decay rate Γ_i is

$$\Gamma_i = \frac{S}{2\hbar m_1} \int |M|^2 \, (2\pi)^4 \delta^4(P_1 - P_2 - P_3 - ... - P_n) \prod_{j=2}^{n} 2\pi \, \delta(P_j^2 - m_j^2 c^2) \, \Theta(P_j^0) \, \frac{d^4 P_j}{(2\pi)^4} \tag{2.2}$$

where all the momenta P are 4-momenta.



For the collision version of the golden rule it is assumed that two particles with mass m_1 and m_2 come in and result in the process of the collision for channel i in particles with mass m_3, ... , m_n which are going out. Two particles colliding is the easiest case, and it is difficult to let three particles collide at the same time, and even then it is possible to build the three-particle collision out of two-particle collisions. The particles going out could be the same as the particles coming in, but the particles coming out and their number could be different.

The collision process for channel i is therefore $m_1 + m_2 \to m_3 + m_4 + ... + m_n$. The corresponding scattering cross-section σ_i is

$$\sigma_i = \frac{S\hbar^2}{4\sqrt{(P_1 P_2)^2 - (m_1 m_2 c^2)^2}} \int |M|^2 \, (2\pi)^4 \delta^4(P_1 + P_2 - P_3 - ... - P_n) \prod_{j=3}^{n} 2\pi \, \delta(P_j^2 - m_j^2 c^2) \, \Theta(P_j^0) \, \frac{d^4 P_j}{(2\pi)^4} \tag{2.3}$$

where all the momenta P are 4-momenta.

There are a lot of common ingredients between (2.2) and (2.3). The factor S in front of the two integrals is $S = \frac{1}{S_1!} \cdot \frac{1}{S_2!} \cdot ...$ where S_i is the number of identical outgoing particles of type i. If, for example, two electrons come out of a process the integral will count both of them twice once as m_i and once as m_j although they are indistinguishable. The case where one electron has number i and the other number j is the same as the case where the two numbers are switched. It should only be counted once.

Both formulas carry a factor $|M|^2$ which is the quantum mechanical amplitude. It is called the *dynamical information*, and it controls the dynamics and connects the things coming in with the things going out as indicated in the two diagrams above, but this part of the process cannot be observed. (Note that in all the diagrams including the Feynman diagrams time flows from left to right as a convention and not from bottom to top as sometimes used by an alternative convention.) The rest of the two expressions shows how the kinematic quantities associated with the incoming particles are related to the kinematic quantities of the outgoing particles.

The factor $(2\pi)^2\delta^4(P...P)$ enforces overall 4-momentum conservation because it makes sure that the sum of the incoming 4-momenta is equal to the sum of the outgoing 4-momenta. (The delta function $\delta(x - a)$ is zero everywhere except for $x = a$. It is especially useful in integrals where it acts as in $\int f(x)\delta(x - a)\, dx = f(a)$.)

The expression $\delta(P_j^2 - m_j^2 c^2)$ also with a delta function forces the outgoing particles to satisfy the mass-shell condition. It is placed within a product over all outgoing particles. One single factor written with indices is $\delta(P_{j\,\mu} P_j^\mu - m_j^2 c^2)$ which makes sure that P_j^2 in any frame is invariant and therefore is equal to $P_j^2 = m_j^2 c^2$ in the rest frame. This part of the two golden rules ensures that the outgoing particles are real. Energy and momentum cannot be arbitrarily distributed to the outgoing particles but each of them must satisfy the mass-shell condition.

The factor $\Theta(P_j^0)$ with the step function enforces $E > 0$ for all outgoing particles. One could put in a negative energy in a 4-momentum and satisfy the mass-shell condition. Thus, the mass-shell condition alone does not prevent negative energies for the outgoing particles.

The integral goes over all 4-momenta from minus infinity to plus infinity. Therefore the mass-shell condition and $E > 0$ are needed to eliminate non-physical 4-momenta from the integral. The spatial momenta can be positive or negative but the energy cannot because it is kinetic energy in the situations here.

The golden rule simply states that − dynamics aside − all kinematic configurations consistent with 4-momentum conservation, positive energy, and mass-shell condition are equally likely. So the more final configurations that are compatible there are or, in other words, the larger the phase space is for the outgoing particles, the larger is the probability. The integral is over all possible combinations of outgoing energy and momentum which are consistent with conservation principles and reality constraints.

At this point one usually cannot go further since $|M|^2$ will often depend on P_j and is therefore needed before one can integrate. It will turn out that $|M|^2$ itself contains integrals and there is no guarantee that the expressions contain enough delta functions to take care of all the integrals. However there are some special situations where the kinematics is so tightly constrained that one can go a bit further. First, one can always break up $d^4P_j = dP^0\, d\vec{P}_j$ and use $\delta(P_j^2 - m_j^2c^2) = \delta((P_j^0)^2 - \vec{P}^2 - m_j^2c^2)$ to perform the dP_j^0 integral using the properties that $\delta(x^2-k^2) = \frac{1}{2k}[\delta(x-k)+\delta(x+k)]$ for a constant $k>0$. This gives

$$\Gamma_i = \frac{S}{2\hbar m_1} \int |M|^2 (2\pi)^4 \delta^4(P_1 - P_2 - P_3 - ... - P_n) \prod_{j=2}^{n} \frac{1}{2\sqrt{\vec{P}_j^2 + m_j^2c^2}} \frac{d^3\vec{P}_j}{(2\pi)^3}$$

$$\sigma_i = \frac{S\hbar^2}{4\sqrt{(P_1P_2)^2 - (m_1m_2c^2)^2}} \int |M|^2 (2\pi)^4 \delta^4(P_1 + P_2 - P_3 - ... - P_n) \prod_{j=3}^{n} \frac{1}{2\sqrt{\vec{P}_j^2 + m_j^2c^2}} \frac{d^3\vec{P}_j}{(2\pi)^3}$$

for the decay rates and scattering cross-sections.

Two cases are so tightly constrained by the kinematics that one has enough delta functions to evaluate all of the integrals without needing the functional form of $|M|^2$. One case is the decay of a particle at rest into two particles. The decay rate for this two-body decay $m_1^{\text{rest}} \to m_2 + m_3$ is

$$\Gamma = \frac{S\left|\vec{P}\right|}{8\pi\hbar m_1^2 c}|M|^2 \qquad \text{with } \left|\vec{P}\right| = \frac{c}{2m_1}\sqrt{m_1^4 + m_2^4 + m_3^4 - 2m_1^2m_2^2 - 2m_1^2m_3^2 - 2m_2^2m_3^2} \tag{2.4}$$

where $|\vec{P}|$ is the magnitude of momentum of either outgoing particle because they are the same due to $\vec{P}_{\text{tot}} = 0$. If one fixes m_1 and plots Γ as a function of m_1, m_3 one will find that it grows with increasing mass difference. Also the two-body scattering $m_1 + m_2 = m_3 + m_4$ in the center of momentum frame simplifies to

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{S|M|^2}{(E_1+E_2)^2} \frac{\left|\vec{P}_f\right|}{\left|\vec{P}_i\right|} \tag{2.5}$$

where $|\vec{P}|_f$ is the magnitude of momentum for either outgoing particle and $|\vec{P}|_i$ is the magnitude of momentum for either incoming particle because $\vec{P}_{\text{tot}} = 0$ in the center of momentum frame. In the two-body decay and the two-body scattering the quantity $|M|^2$ is not needed to calculate the integral but only afterwards to determine the decay rate or the scattering cross-section, respectively. The two cases are so simple because the fact that the two outgoing particles must have the same magnitude of momentum but move in opposite directions restrict the possibilities so strongly that there is not much freedom left.

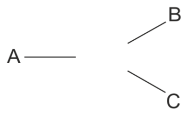
2.3 The Simplified Feynman Rules

The remaining question is how to calculate the quantity M in the form $|M|^2$ used in the golden rules (2.4) and (2.5). As often in physics one does not start with the complete theory but with a simplified toy version. Here the simple toy model Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\Phi_A\partial^\mu\Phi_A - \frac{1}{2}\left(\frac{m_Ac}{\hbar}\right)^2\Phi_A^2 + \frac{1}{2}\partial_\mu\Phi_B\partial^\mu\Phi_B - \frac{1}{2}\left(\frac{m_Bc}{\hbar}\right)^2\Phi_B^2 + \frac{1}{2}\partial_\mu\Phi_C\partial^\mu\Phi_C - \frac{1}{2}\left(\frac{m_Cc}{\hbar}\right)^2\Phi_C^2$$
$$+ g\,\Phi_A\Phi_B\Phi_C$$

is used which is based on the so-called ABC theory without spin. The first line shows three spin-0 fields freely propagating in space and the second line shows how they interact. This interaction term is the simplest possible interaction. The three fields are real and therefore the corresponding particles are their own antiparticle written as $\bar{A} = A$ because for an antiparticle $\bar{A} \neq A$ the field Φ_A needs to be complex and similarly for Φ_B and Φ_C. (This is reflected in Feynman diagrams through lines without arrows. Later when looking at Feynman diagrams of electrons and positrons, arrows pointing forward in time correspond to electrons and arrows pointing backward in time correspond to positrons.) Adding the constraint $m_A > m_B + m_C$ ensures that the particle A can decay into B and C but neither B can decay into A and C nor C can decay into A and B. From the interaction term $g\,\Phi_A\Phi_B\Phi_C$ in the Lagrangian one can conclude that the basic interaction vertex connects A, B and C as the first order contribution shown in figure 2.1 where a A meats a B and a C simultaneously because the interaction term is a product of all three fields. The interaction term just says that three particles meet.

The first step is to draw the diagrams. Beginning with the initial state A and the final state $B + C$ one connects them in all possible ways using the interaction vertices. Note that one can rotate the vertices such that B or C are on the left side, but one cannot change the particle content such that there are two B instead of one B and a C. Rotating a vertex however changes the interpretation.



At first it might seem obvious how to connect the pieces but one quickly realizes that this is not so easy. There are many complicated ways to combine only vertices with A, B, C as shown for first- and third-order diagrams in figure 2.1. The process of adding more and more complicated combinations of such vertices never stops, and the sum in this figure is infinite. Wherever there is a line with A, one can introduce a BC bubble as in the first self-energy-corrections term. All these diagrams are part of the M amplitude for A coming in and B and C coming out.

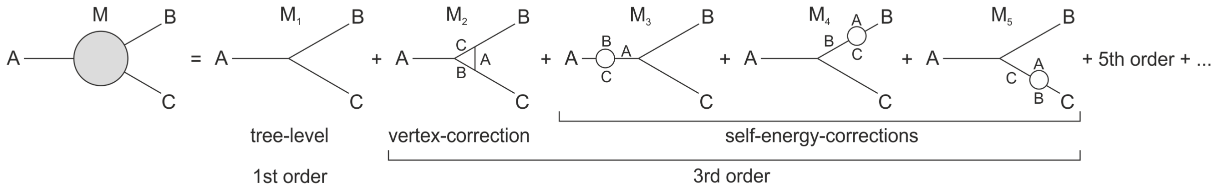


Figure 2.1: Example decay of A into B and C

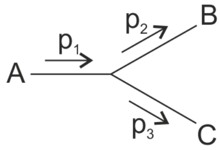
The first diagram has one vertex, the next four diagrams have three vertices, and afterwards would come diagrams with five vertices and so on. The more complex the diagrams are, the more vertices they contain. But for every vertex one has to put in a factor of the coupling constant g. The first-order vertex has a factor g, the four third-order vertices have a factor g^3 and so on. Thus, as long as $g < 1$ the higher-order diagrams contribute less and less. This is in a sense a Taylor expansion in the interaction strength g^n, and it often is enough to only consider the lowest orders.

The first third-order diagram in figure 2.1 looks as if the vertex itself has been blown up into a triangle and is called a vertex correction. The other three third-order diagrams are called self-energy corrections. These diagrams are needed to calculate $|M|^2$, but they are not numbers and have to be converted into numbers first. The rules to do that can be derived from Quantum Field Theory, but here they are just given as facts.

One evaluates each diagram to get M_i using the following rules for the ABC theory which will change slightly for the full Standard Model:

1. Label all momenta with p_i for external momenta (real particles) and q_i for internal momenta (virtual particles) with arrows next to the particle lines. This allows to keep track of momenum flow which is different from particle identity flow. For each p_i the arrow must go forward in time, but the arrows for the q_i can go in one of the two directions or may even happen in an instant of time.
2. For each vertex write a factor of $-ig$ where g is the coupling strength.
3. For each internal line one writes a factor $i/(q_j^2 - m_j^2c^2)$ where $q_j^2 \neq m_j^2c^2$ because the particles are virtual and are therefore not restricted by the mass-shell condition.
4. For each vertex conserve 4-momentum with a factor of $(2\pi)^4\,\delta^4(P_{\text{tot in}} - P_{\text{tot out}})$ where P is the sum over the p_i and q_i.
5. Integrate over all internal momenta with $\int 1/(2\pi)^4 d^4q_j$.
6. After all this, one will have an overall $(2\pi)^4\,\delta^4(P_{\text{tot in}} - P_{\text{tot out}})$ which one should eliminate and replace with a factor i to get M because this term will reappear in Fermi's golden rule and the square of a delta function as for $|M|^2$ is not well-defined.

As a easy example, the first order diagram in figure 2.1 is shown. Since there are no internal momenta, just the three external momenta p_i with $i \in \{1, 2, 3\}$ are assigned to the three lines in the first step. There is only one vertex to which a factor of $-ig$ is added in the second step. The third and fifth step can be skipped because there are no internal momenta. The fourth step adds a factor $(2\pi)^4\, \delta^4(p_1 - p_2 - p_3)$, and the sixth step replaces this factor by a factor i. The result is therefore

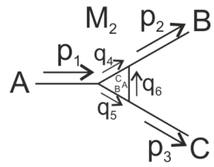


$$-ig(2\pi)^4\, \delta^4(p_1 - p_2 - p_3) \Rightarrow M_1 = g$$

and this is not surprising because nothing interesting happens in this diagram. There is a single interaction and its strength is g.

To get M, also the contributions for third-order diagrams, fifth-order diagrams and so on are needed to be added to M, and finally $|M|^2$ has to be calculated for Fermi's golden rule. Only the third-order contributions are calculated here as examples for the application of the Feynman rules.

The vertex-correction diagram gives



$$\iiint (-ig)^3 \frac{i}{q_4^2 - m_C^2 c^2} \cdot \frac{i}{q_5^2 - m_B^2 c^2} \cdot \frac{i}{q_6^2 - m_A^2 c^2} \cdot (2\pi)^4 \delta^4(p_1 - q_4 - q_5)(2\pi)^4 \delta^4(q_4 + q_6 - p_2) \cdot (2\pi)^4 \delta^4(q_5 - q_6 - p_3) \frac{d^4 q_4}{(2\pi)^4} \frac{d^4 q_5}{(2\pi)^4} \frac{d^4 q_6}{(2\pi)^4}$$

after the fifth step. One can use $\delta^4(p_1 - q^4 - q^5)$ to do the q_4 integral and gets

$$\iint (-ig)^3 \frac{i}{(p_1 - q_5)^2 - m_C^2 c^2} \cdot \frac{i}{q_5^2 - m_B^2 c^2} \cdot \frac{i}{q_6^2 - m_A^2 c^2} \cdot (2\pi)^4 \delta^4(p_1 - q_5 + q_6 - p_2) \cdot (2\pi)^4 \delta^4(q_5 - q_6 - p_3) \frac{d^4 q_5}{(2\pi)^4} \frac{d^4 q_6}{(2\pi)^4}$$

by replacing $q_4 = p_1 - q_5$ everywhere. Using similarly $\delta^4(q_5 - q_6 - p_3)$ for q_6 gives

$$\int (-ig)^3 \frac{i}{(p_1 - q_5)^2 - m_C^2 c^2} \cdot \frac{i}{q_5^2 - m_B^2 c^2} \cdot \frac{i}{(q_5 - p_3)^2 - m_A^2 c^2} \cdot (2\pi)^4 \delta^4(p_1 - p_2 - p_3) \frac{d^4 q_5}{(2\pi)^4}$$

or

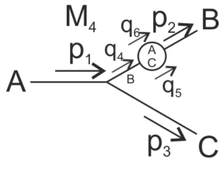
$$(-ig)^3 i^3 (2\pi)^4 \delta^4(p_1 - p_2 - p_3) \int \frac{1}{(p_1 - q_5)^2 - m_C^2 c^2} \cdot \frac{1}{q_5^2 - m_B^2 c^2} \cdot \frac{1}{(q_5 - p_3)^2 - m_A^2 c^2} \frac{d^4 q_5}{(2\pi)^4}$$

resulting in

$$M_2 = (-ig)^3 i^4 \int \frac{1}{(p_1 - q_5)^2 - m_C^2 c^2} \cdot \frac{1}{q_5^2 - m_B^2 c^2} \cdot \frac{1}{(q_5 - p_3)^2 - m_A^2 c^2} \frac{d^4 q_5}{(2\pi)^4}$$

after the sixth step.

Similarly one of the self-energy-correction diagrams gives



$$\iiint (-ig)^3 \frac{i}{q_4^2 - m_B^2 c^2} \cdot \frac{i}{q_5^2 - m_C^2 c^2} \cdot \frac{i}{q_6^2 - m_A^2 c^2} \cdot (2\pi)^4 \delta^4(p_1 - q_4 - p_3)(2\pi)^4 \delta^4(q_4 - q_5 - q_6) \cdot (2\pi)^4 \delta^4(q_5 + q_6 - p_2) \frac{d^4 q_4}{(2\pi)^4} \frac{d^4 q_5}{(2\pi)^4} \frac{d^4 q_6}{(2\pi)^4}$$

after the fifth step and results in

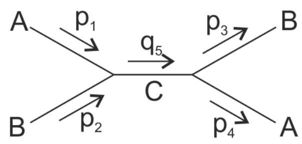
$$M_4 = (-ig)^3 i^4 \frac{1}{p_2^2 - m_B^2 c^2} \int \frac{1}{q^2 - m_C^2 c^2} \cdot \frac{1}{(p_2 - q)^2 - m_A^2 c^2} \frac{d^4 q}{(2\pi)^4}$$

after the sixth step.

The other two self-energy-correction diagrams can be calculated the same way. However, to get a more precise value for M also the higher-order diagrams must be taken into account.

2.4 Scattering Examples

As a first example for a scattering process, M is calculated for the lowest order of $A+B \to A+B$. There are two diagrams.



The first five steps for the first diagram give the integral

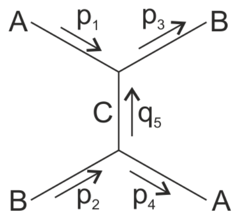
$$\int (-ig)^2 \frac{i}{q_5^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_1 + p_2 - q_5) \cdot (2\pi)^4 \delta^4(q_5 - p_3 - p_4) \frac{d^4 q_5}{(2\pi)^4}$$

$$= (-ig)^2 i \frac{1}{(p_3 + p_4)^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4)$$

using $q_5 = p_3 + p_4$. The result is

$$M_1 = \frac{g^2}{(p_3 + p_4)^2 - m_C^2 c^2}$$

after the sixth step.



The first five steps for the second diagram give the integral

$$\int (-ig)^2 \frac{i}{q_5^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_1 + q_5 - p_3) \cdot (2\pi)^4 \delta^4(p_2 - q_5 - p_4) \frac{d^4 q_5}{(2\pi)^4}$$

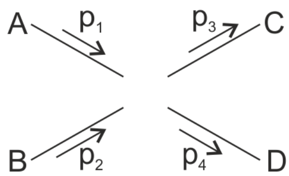
$$= (-ig)^2 i \frac{1}{(p_2 - p_4)^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4)$$

using $q_5 = p_2 - p_4$. The result for both diagrams is

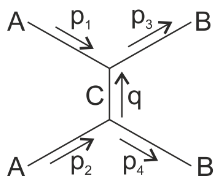
$$M_1 + M_2 = g^2 \left[\frac{1}{(p_3 + p_4)^2 - m_C^2 c^2} + \frac{1}{(p_2 - p_4)^2 - m_C^2 c^2} \right]$$

after the sixth step.

The interpretation of the first diagram shows that the particles A and B disappear for some time and there is only a C. After some time the C disappears and an A and a B emerge. In the second diagram there is always an A and a B there. That has kinematic consequences. The momentum of particle C must be the total momentum of A and B in the first diagram, while the momentum of particle C is the difference in the second diagram because it is the amount of momentum A gives to B and vice versa. Kinematic considerations help figuring out whether two diagrams are the same or not.



When evaluating possible Feynman diagrams in order to add several contributions it helps to start with all of the external labels. Then one can fill in the rest of the diagram. This way one can avoid mistakenly using a momentum p_i for different particles in different diagrams. It is important when adding the contributions M_i to M that a single external momentum label always refers to the same particle.

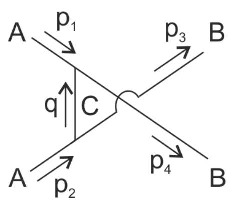


The question is whether the two diagrams on the right side for a second scattering example $A + A \to B + B$ are different. The first diagram results in

$$\int (-ig)^2 \frac{i}{q^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_1 + q - p_3)(2\pi)^4 \delta^4(p_2 - q - p_4) \frac{d^4 q}{(2\pi)^4} =$$

$$(-ig)^2 \frac{i}{(p_3 - p_1)^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_2 - p_3 + p_1 - p_4) \Rightarrow M = \frac{g^2}{(p_3 - p_1)^2 - m_C^2 c^2}$$

and the second diagram gives

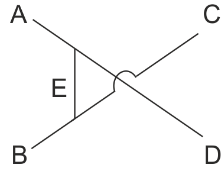


$$\int (-ig)^2 \frac{i}{q^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_2 - q - p_3)(2\pi)^4 \delta^4(p_1 + q - p_4) \frac{d^4 q}{(2\pi)^4} =$$

$$(-ig)^2 \frac{i}{(p_4 - p_1)^2 - m_C^2 c^2} (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \Rightarrow M = \frac{g^2}{(p_4 - p_1)^2 - m_C^2 c^2}$$

which shows that they are different. One can use $p_1 = 100$, $p_2 = 2$, $p_3 = 101$, $p_4 = 1$ and gets $q = 1$ in the first diagram and $q = -99$ in the second. Thus one should both include as long as they are allowed by the rules of the theory.

To show that two lines in a diagram do not intersect one can draw a little hop over one of the lines. This is useful because if one tries to figure out all possible diagrams it helps to draw all incoming and outgoing lines always in the same place. It is still not always easy to see whether two diagrams are equal or different.Feynman diagrams stay the same when vertices are moved around continuously. Therefore there is some sense for topology and a bit of experience needed when working with Feynman diagrams.



For a scattering process where two particles come in and two particles go out $\frac{d\sigma}{d\Omega}$ is given by (2.5) in the center of momentum frame, and for a decay of one particle at rest into two particles Γ is given by (2.4). Calculating $|M|^2$ as well as Γ or $\frac{d\sigma}{d\Omega}$ is a tedious job as the above examples already show. To calculate the lifetime of a particle, one has to calculate Γ for all possible channels because lifetime does not depend on what a particle decays into.

2.5 Feynman Rules from a Lagrangian

The ABC-theory had a simple Lagrangian with a Klein-Gorden term $\frac{1}{2}\partial_\mu\Phi_A\partial^\mu\Phi_A - \frac{1}{2}\left(\frac{m_A c}{\hbar}\right)^2\Phi_A^2$ for Φ_A (and similarly for the other two fields Φ_B, Φ_C) and an interaction term $-g\Phi_A\Phi_B\Phi_C$. In evaluating diagrams built from lines labeled A, B or C with one line of each label joining in a vertex, the vertex factors $-ig$ and virtual particle propagators $i/(q_j^2 - m_j^2c^2)$ have been used.

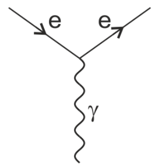
This is actually a systematic way to extract the Feynman rules from a given Lagrangian but Quantum Field Theory is required to derive it. The results can be stated simply. For the vertex factor, one writes $i\mathcal{L}_{\text{int}}$ for an interaction term translated with $i\hbar\partial_\mu \to P_\mu$ or $\partial_\mu \to -\frac{i}{\hbar}P_\mu$ into momentum space (giving $-ig\Phi_A\Phi_B\Phi_C$ in the ABC example), and afterwards one erases the field variables (resulting in $-ig$ in the ABC example). For the propagators, one writes the relevant free-particle equation of motion in momentum space, for example, the Klein-Gordon version $\partial_\mu\partial^\mu\Phi + (\frac{mc}{\hbar})^2\Phi = 0$ which is $[P^2-(mc)^2]\Phi = 0$ in momentum space, erases overall factor of $\hbar$, and multiplies the inverse of the term in brackets times i giving $i/[P^2 - (mc)^2]$.

3 Application to Quantum Electrodynamics

3.1 Vertex Factor and Propagators

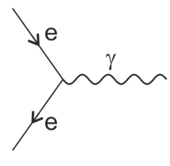
The Lagrangian for Quantum Electrodynamics is

$$\mathcal{L}_{\text{QED}} = i\hbar c\bar{\Psi}\gamma^\mu\partial_\mu\Psi - mc^2\bar{\Psi}\Psi - q\bar{\Psi}\gamma^\mu\Psi A_\mu - \frac{1}{16\pi}\left(\partial^\mu A^\nu - \partial^\nu A^\mu\right)\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right)$$



and the interaction term is $-q\bar{\Psi}\gamma^\mu\Psi A_\mu$ with a spinor $\bar{\Psi}$, a spinor Ψ and a gauge field A_μ. The fundamental vertex will therefore have three lines coming together. Because the gauge field is different than the other two spinor fields, it will get a squiggly line. This Lagrangian covers real physics and it is known that Ψ describes a charged fermion such as the electron which is different than its antiparticle positron. This means that one has to assign arrows to the lines. An arrow pointing in the direction of time represents a particle, and an arrow pointing opposite to the direction of time represents an antiparticle. (The difference between Ψ and $\bar{\Psi}$ is not the difference between a particle and an antiparticle but is just mathematically needed to get a scalar when Ψ and $\bar{\Psi}$ are multiplied.)

The fundamental vertex is a squiggly line for the gauge field γ and two arrows for the spinor fields. Because the photon γ is its own antiparticle, there is no need to draw an arrow on the squiggly line. This is different for the charged particle. Depending how the fundamental vertex is rotated, the Feynman diagram shows different physical situations. An electron may emit a photon and continue as an electron. In this situation both arrows on the lines for the electrons point in the direction in which the time flows. The same fundamental vertex rotated shows an electron meeting a positron such that the two particles annihilate into a photon. In this



situation the arrow for the electron points in the direction of time and the arrow for the positron points in the opposite direction. Variations of this fundamental vertex must satisfy charge conservation.

The fundamental vertex is the only vertex in Quantum Electrodynamics. One can vary the topology as long as charge conservation is not violated, and one can replace the electron by another charged particle. The vertex looks the same for an electron or, for example, a muon.

The vertex factor one gets from $i\mathcal{L}_{\text{int}} = -iq\bar{\Psi}\gamma^\mu\Psi A_\mu$ by removing $\bar{\Psi}$, Ψ and A_μ in the form $\sqrt{(\hbar c)/(4\pi)}A_\mu$ due to units carried by A_μ is

$$-i\sqrt{\frac{4\pi}{\hbar c}}q\gamma^\mu = ig_e\gamma^\mu \qquad \text{with} \qquad g_e = -\sqrt{\frac{4\pi}{\hbar c}}q$$

which is more difficult than the vertex factor $-ig$ in the ABC theory because it contains γ^μ.

For the propagator there can be virtual photos but also virtual electrons and positrons. For the charged particles the Dirac equation of motion $i\gamma^\mu\partial_\mu\Psi - \frac{mc}{\hbar}\Psi = 0$ which is $[\gamma^\mu P_\mu - mc]\Psi = 0$ in momentum space must be used leading to

$$\frac{i(\gamma^\mu P_\mu + mc)}{P^2 - m^2c^2}$$

because the quantity $\gamma^\mu P_\mu - mcI$ is a matrix in spin space and has to be turned into a matrix inverse. (Note that in spacetime the quantities $\gamma^\mu P_\mu$ and mc are scalars.) For the photon one starts with the massive Proca equation $\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) + \left(\frac{mc}{\hbar}\right)^2 A^\nu = 0$ because massive gauge fields are needed for the weak interactions later on. This becomes $-P_\mu P^\mu A^\nu + P_\mu P^\nu A^\mu + (mc)^2 A^\nu$ in momentum space by replacing ∂_μ by $-\frac{i}{\hbar}P_\mu$ and simplifying the factors $\frac{i}{\hbar}$. To get something of the form $[...]A^\nu$ some steps are needed. By multiplying with $\eta_{\lambda\nu}$ one gets

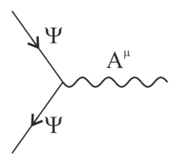
$$\begin{aligned}
\eta_{\lambda\nu}(-P_\mu P^\mu A^\nu + P_\mu P^\nu A^\mu + (mc)^2 A^\nu) &= \eta_{\lambda\nu}\left(-P^2 A^\nu + P_\mu P^\nu A^\mu\right) + (mc)^2\eta_{\lambda\nu}A^\nu \\
&= \eta_{\lambda\nu}\left(-P^2 + (mc)^2\right)A^\nu + P_\mu P_\lambda A^\mu = \eta_{\lambda\nu}\left(-P^2 + (mc)^2\right)A^\nu + P_\nu P_\lambda A^\nu \\
&= \left[\eta_{\lambda\nu}\left(-P^2 + (mc)^2\right) + P_\nu P_\lambda\right]A^\nu = \left[\eta_{\mu\nu}\left(-P^2 + (mc)^2\right) + P_\mu P_\nu\right]A^\nu
\end{aligned}$$

after renaming some indices consistently. The quantity in brackets is a $(0,2)$ tensor in spacetime, but is trivial in spin space. The inverse of $[\eta_{\mu\nu}(-P^2 + (mc)^2) + P_\mu P_\nu]$ gives the propagator for the gauge fields

$$i\left[\eta_{\mu\nu}\left(-P^2 + (mc)^2\right) + P_\mu P_\nu\right]^{-1} = \begin{cases} \dfrac{-i}{P^2 - (mc)^2}\left[\eta_{\mu\nu} - \dfrac{P_\mu P_\nu}{(mc)^2}\right] & m \neq 0 \\[2ex] \dfrac{-i\eta_{\mu\nu}}{P^2} & m = 0 \end{cases} \tag{3.1}$$

where the case $m = 0$ is the propagator for the photon and the case $m \neq 0$ will be useful later for the weak interactions. (Note that $\eta_{\mu\nu}^{-1} = \eta_{\mu\nu}$ and that for $m = 0$ due to transversality $P_\nu A^\nu = 0$ is satisfied leaving $-\eta_{\mu\nu}P^2A^\nu = 0$. The step from $m \neq 0$ to $m = 0$ is a discontinuous limit because the number of polarization states changes discretely from 3 to 2.)

3.2 Useful Equations in Quantum Electrodynamics



The fundamental diagram from which all diagrams in Quantum Electrodynamics are build has two lines with arrows labeled Ψ for the charged fermions and one squiggly line labeled A^μ for the gauge field. Restricting the theory to charged leptons and photons, Ψ can be an electron e, a muon μ or a tauon τ, and A^μ is a photon γ.

In the ABC theory, all degrees of freedom were scalars so the order was unimportant and M was guaranteed to be a scalar. In Quantum Electrodynamics on the other side the three charged leptons are spinors and the photon is a vector such that the order is important and one has to be careful to ensure a scalar M. For a scalar field the only two important quantities needed are the direction it moves and the energy it has. Spinors have in addition a spin degree of freedom. The complications have to do with the question what the spin is doing.

Useful expressions:

	Electron	Positron	Photon
Wavefunction with $s \in \{1,2\}$	$\Psi(X) = a\, e^{-\frac{i}{\hbar}P\cdot X}\, u^{(s)}(P)$	$\Psi(X) = a\, e^{\frac{i}{\hbar}P\cdot X}\, v^{(s)}(P)$	$A_\mu(X) = a\, e^{-\frac{i}{\hbar}P\cdot X}\epsilon_\mu^{(s)}$
Equ. of motion (in P space)	$(\gamma^\mu P_\mu - mc)u = 0$ (Dirac equation)	$(\gamma^\mu P_\mu + mc)u = 0$ (Dirac equation)	$P^\mu \epsilon_\mu = 0$ (transversality) $\epsilon^0 = 0$ (gauge choice)
Adjoint	$\bar{u} = u^\dagger \gamma^0$	$\bar{v} = v^\dagger \gamma^0$	$\epsilon^{\mu *}$
Adjoint equ. of motion	$\bar{u}(\gamma^\mu P_\mu - mc) = 0$	$\bar{v}(\gamma^\mu P_\mu + mc) = 0$	$P_\mu \epsilon^{\mu *} = 0$
Orthonormality	$\bar{u}^{(s)}u^{(s')} = 2mc\,\delta_{ss'}$	$\bar{v}^{(s)}v^{(s')} = -2mc\,\delta_{ss'}$	$\epsilon_\mu^{(s)}\epsilon^{(s')\mu^*} = -\delta_{ss'}$
Completeness	$\sum_s u^{(s)}\bar{u}^{(s)} = \gamma^\mu P_\mu + mc$	$\sum_s v^{(s)}\bar{v}^{(s)} = \gamma^\mu P_\mu - mc$	$\sum_s \epsilon_i^{(s)}\epsilon_j^{(s)^*} = \delta_{ij} - \hat{P}_i\hat{P}_j$

where $u^{(s)}$ and $v^{(s)}$ are the Dirac spinors with s labeling the two different spin states, $\epsilon^{(s)}$ are the constant polarization tensor with s labeling the two different polarization states, and $\hat{P}_i$ are the spatial components of the polarization vector. (Note that $\bar{u}^{(s)}u^{(s')}$ is a scalar while $\sum_s u^{(s)}\bar{u}^{(s)}$ is a matrix in spin space.)

One new complication is that in most experiments unpolarized incoming states have been used and it has been summed over all outgoing states because one does not care about the spin polarization. This means in calculating $|M|^2$ one has to average over incoming spin states and sum over outgoing spin states. This is where the orthonormality and completeness equations will turn out to be useful.

The quantities $u^{(s)}$ and $v^{(s)}$ are four-component spinors and are plane-wave solutions to the Dirac equation. They are

$$
\begin{aligned}
u^{(1)} &= \sqrt{\frac{E+mc^2}{c}}\begin{pmatrix} 1 \\ 0 \\ \frac{cP_z}{E+mc^2} \\ \frac{c(P_x+iP_y)}{E+mc^2} \end{pmatrix} & u^{(2)} &= \sqrt{\frac{E+mc^2}{c}}\begin{pmatrix} 0 \\ 1 \\ \frac{c(P_x-iP_y)}{E+mc^2} \\ \frac{-cP_z}{E+mc^2} \end{pmatrix} \\
v^{(1)} &= \sqrt{\frac{E+mc^2}{c}}\begin{pmatrix} \frac{c(P_x-iP_y)}{E+mc^2} \\ \frac{-cP_z}{E+mc^2} \\ 0 \\ 1 \end{pmatrix} & v^{(2)} &= -\sqrt{\frac{E+mc^2}{c}}\begin{pmatrix} \frac{cP_z}{E+mc^2} \\ \frac{c(P_x+iP_y)}{E+mc^2} \\ 1 \\ 0 \end{pmatrix}
\end{aligned}
\tag{3.2}
$$

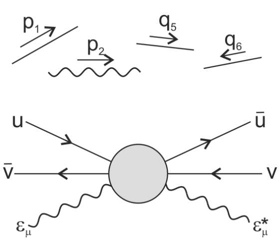
in the conventions used here. The spinor matrices are

$$
\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \qquad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
$$

where the σ matrices are the Pauli matrices and the γ matrices are the Dirac matrices.

3.3 Feynman Rules for Quantum Electrodynamics

The Feynman rules adapted for Quantum Electrodynamics are:

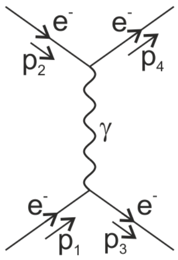


1. Draw diagram and label matter (fermion) lines with arrows to distinguish particles from antiparticles. Label momenta for external lines with p_i aligned along time direction and for internal lines with q_j.
2. Each external line gets a factor according to its type. An incoming matter particle gets a u, and an outgoing matter particle gets a $\bar{u}$. An incoming matter antiparticle gets a $\bar{v}$, and an outgoing matter antiparticle gets a v. An incoming photon gets an ϵ_μ, and an outgoing photon gets an ϵ^*_μ. Number the factor with momenta such as $u(3)$ for the incoming line labeled with momentum p_3. (Note that each factor is either a 4-spinor for matter or a 4-vector for photons.)
3. Each vertex gets a factor $ig_e\gamma^\mu$ where $g_e = e\sqrt{\frac{4\pi}{\hbar c}}$. Overall spin matrix.

4. Each internal line gets a propagator factor:
   Matter particle or antiparticle: $\frac{i(\gamma^\mu q_\mu + mc)}{q^2 - m^2c^2}$ Overall spin matrix.
   Photon: $-\frac{i\eta_{\mu\nu}}{q^2}$ Overall tensor (μ, ν match vertex indices)
5. Conserve 4-momentum at each vertex with $(2\pi)^4\,\delta^4(p_{\text{in}}+q_{\text{in}}-p_{\text{out}}-q_{\text{out}})$.
6. Integrate over each internal momentum q with $\int \frac{d^4q}{(2\pi)^4}$.
7. Cancel overall $(2\pi)^4\,\delta^4(p_{\text{tot in}} - p_{\text{tot out}})$ and multiply by i to get M.
8. Antisymmetrize between diagrams related by switching
   - two incoming electrons/positrons
   - two outgoing electrons/positrons
   - one incoming electron/positron with one outgoing positron/electron
9. Get the order right.

Because here index notation is not used for spinors and spin matrices but matrix multiplication, one has to get the order right as requested in the last point. (For spacetime indices it is not as important.) Thus, in order to get the right order for spinor elements one should make "spinor sandwiches" from matter lines by starting with an outgoing matter particle line ($\bar{u}$) or incoming antimatter line ($\bar{v}$) and following along only matter segments writing vertex factors and internal matter propagators as one encounters them, until one eventually emerge on an outgoing antimatter (v) or incoming matter (u) line. The photon factors are less tricky since index notation gets them right.

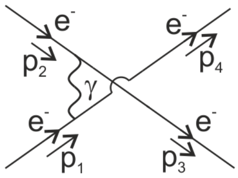
As an example electron-electron scattering $e^- + e^- \to e^- + e^-$ is shown:



$$\Rightarrow M_1 = i\bar{u}(3)ig_e\gamma^\mu u(1)\,\bar{u}(4)ig_e\gamma^\nu u(2)\left(\frac{-i\,\eta_{\mu\nu}}{(p_4-p_2)^2}\right)$$

because

$$\int \bar{u}(3)ig_e\gamma^\mu u(1)\,\bar{u}(4)ig_e\gamma^\nu u(2)\left(\frac{-i\,\eta_{\mu\nu}}{q^2}\right)(2\pi)^4\delta^4(p_1-q-p_3)$$
$$(2\pi)^4\delta^4(p_2+q-p_4)\,\frac{d^4q}{(2\pi)^4}$$
$$= \bar{u}(3)ig_e\gamma^\mu u(1)\,\bar{u}(4)ig_e\gamma^\nu u(2)\left(\frac{-i\,\eta_{\mu\nu}}{(p_4-p_2)^2}\right)(2\pi)^4\delta^4(p_1+p_2-p_3-p_4)$$



There is another second-order diagram which can be handled similarly with $3 \leftrightarrow 4$:

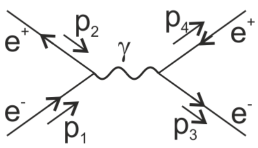
$$\Rightarrow M_2 = i\bar{u}(4)ig_e\gamma^\mu u(1)\,\bar{u}(3)ig_e\gamma^\nu u(2)\left(\frac{-i\,\eta_{\mu\nu}}{(p_3-p_2)^2}\right)$$

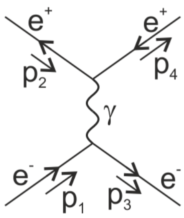
By step 8: $M = M_1 - M_2$

There are no more second-order diagrams but there are obviously additional higher-order diagrams not further explored here.

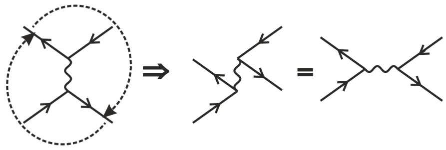
Note that $\bar{u}(3)ig_e\gamma^\mu u(1)$ and similarly $\bar{u}(4)ig_e\gamma^\nu u(2)$ are scalars because $\bar{u}$ is a row vector, γ^μ is a matrix, and u is a column vector all in spin space. Note also that $\eta_{\mu\nu}$ takes care of the factors γ^μ and γ^ν.

When studying electron-muon scattering $e^- + e^- \to \mu^- + \mu^-$, the first diagram is allowed where the lower two electrons are replaced by two muons. The second diagram, however, is not possible because an electron cannot change into a muon in Quantum Electrodynamics while two electrons are indistinguishable and can swap their position.





The case of positron-electron scattering $e^+ + e^- \to e^+ + e^-$ leads to the question how to apply the antisymmetrizing step in the case of the two diagrams where in one diagram an electron and a positron annihilate, emit a photon, and the photon decays into an electron and a positron, while in the other diagram a positron and an electron exchange a photon.

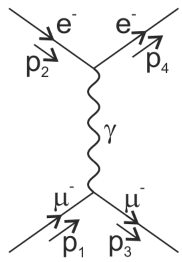


If one diagram gets M_1 and the other M_2, is $M = M_1 + M_2$ or $M = M_1 - M_2$? The right antisymmetrizing is

$$M = M_1 - M_2$$

because switching one outgoing e^- with one incoming e^+ in the second diagram creates a diagram with is topologically equivalent to the first diagram.

Considering electron-muon scattering $e^- + \mu^- \to e^- + \mu^-$ because there is only one diagram of lowest order gives



$$M = i\bar{u}(3) i g_e \gamma^\mu u(1)\, \bar{u}(4) i g_e \gamma^\nu u(2) \left(\frac{-i\, \eta_{\mu\nu}}{(p_4 - p_2)^2} \right)$$

as calculated above. The values for u secretly have all a spin label and one can evaluate M given

$$u(i) = \alpha_i u^{(1)}(i) + \beta_i u^{(2)}(i)$$

for $i \in \{1, 2, 3, 4\}$. One can specify exactly what one puts into the experiment in terms of spin and one can specify what one expects to come out of the experiment but this is not done. In practice one sends in unpolarized beams and does not fix the spin going in, and when one measures what is coming out one measures everything independent of the spin. Thus, for practical calculations one wants to average over all incoming spins (here α_1, β_1, α_2, β_2) and sum over the outgoing spins (here α_3, β_3, α_4, β_4).

This means for $|M|^2$ because M always has factors of the form $\bar{u}(a)\Gamma_1 u(b)$ called spinor sandwiches that there are factors $[\bar{u}(a)\Gamma_1 u(b)][\bar{u}(a)\Gamma_2 u(b)]^*$. The quantities Γ_i are some spin matrices with indices to distinguish the different cases. (Note that if Γ_1 carries spacetime indices μ, ν then Γ_2 will need different indices to avoid confusing contractions.) Because $[\bar{u}(a)\Gamma_2 u(b)]^*$ is a scalar in spin space such that transposing has no effect, it is the same as $[\bar{u}(a)\Gamma_2 u(b)]^\dagger$. This is $u(b)^\dagger \Gamma_2^\dagger \bar{u}(a)^\dagger = u(b)^\dagger \Gamma_2^\dagger (u(a)^\dagger \gamma^0)^\dagger$ or $u(b)^\dagger \Gamma_2^\dagger \gamma^{0\dagger} u(a)$. Using $\gamma^0\gamma^0 = I$ and $\gamma^0 = \gamma^{0\dagger}$ gives $u(b)^\dagger \gamma^0\gamma^0 \Gamma_2^\dagger \gamma^0 u(a) = \bar{u}(b)\bar{\Gamma}_2 u(a)$ with $\bar{\Gamma}_2 = \gamma^0 \Gamma_2^\dagger \gamma^0$ as a definition. Thus, $[\bar{u}(a)\Gamma_1 u(b)][\bar{u}(a)\Gamma_2 u(b)]^* = \bar{u}(a)\Gamma_1 u(b)\bar{u}(b)\bar{\Gamma}_2 u(a)$.

Adding all these factors in $|M|^2$, using the completeness property $\sum_s u^{(s)}\bar{u}^{(s)} = \gamma^\mu P_\mu + mc$ and defining $\not{P} = \gamma^\mu P_\mu$ gives

$$\sum_{s_b} |M|^2 = ...\bar{u}(a)\Gamma_1 u(b)\bar{u}(b)\bar{\Gamma}_2 u(a)... = ...\bar{u}(a)\Gamma_1(\not{p}_b + m_b c)\bar{\Gamma}_2 u(a)...$$

where $\bar{u}(a)$ and $u(a)$ are not next to each other as required for a second application of the completeness property, but because $\bar{u}(a)$ is a row vector, $\Gamma_1 u(b)(\not{p}_b + m_b c)\bar{\Gamma}_2$ is a matrix and $u(a)$ is a column vector the result is a scalar this is the same as the trace of the matrix times the column vector times the row vector. This gives

$$\sum_{s_b} |M|^2 = ...\ \mathrm{Tr}\left[\Gamma_1(\not{p}_b + m_b c)\bar{\Gamma}_2 u(a)\bar{u}(a)\right]...$$

$$\sum_{s_a s_b} |M|^2 = ...\ \mathrm{Tr}\left[\Gamma_1(\not{p}_b + m_b c)\bar{\Gamma}_2(\not{p}_a + m_a c)\right]...$$

where spin matrices occur, but no spinors are left.

This gives a rule of thumb:
To impose sum over s_a, s_b replace $[\bar{u}(a)\Gamma_1 u(b)][\bar{u}(a)\Gamma_2 u(b)]^*$ by $\mathrm{Tr}\Big[\Gamma_1(\not{p}_b + m_b c)\bar{\Gamma}_2(\not{p}_a + m_a c)\Big]$.

Applied to the electron-muon scattering $e^- + \mu^- \to e^- + \mu^-$ with

$$M = i\bar{u}(3) i g_e \gamma^\mu u(1)\, \bar{u}(4) i g_e \gamma^\nu u(2) \left(\frac{-i\, \eta_{\mu\nu}}{(p_4 - p_2)^2} \right) = -\frac{g_e^2}{(p_4 - p_2)^2} \bar{u}(3)\gamma^\mu u(1)\, \bar{u}(4)\gamma^\nu u(2)\, \eta_{\mu\nu} \qquad (3.3)$$

gives

$$|M|^2 = \frac{g_e^4}{(p_4 - p_2)^4} \left[\bar{u}(3) i g_e \gamma^\mu u(1)\, \bar{u}(4) i g_e \gamma^\nu u(2) \eta_{\mu\nu}\right] \left[\bar{u}(3) i g_e \gamma^\kappa u(1)\, \bar{u}(4) i g_e \gamma^\lambda u(2) \eta_{\kappa\lambda}\right]^*$$

$$= \frac{g_e^4}{(p_4 - p_2)^4}\, \mathrm{Tr}\left[\gamma^\mu(\not{p}_1 + m_1 c)\bar{\gamma}^\kappa(\not{p}_3 + m_3 c)\right] \mathrm{Tr}\left[\gamma^\nu(\not{p}_2 + m_2 c)\bar{\gamma}^\lambda(\not{p}_4 + m_4 c)\right] \eta_{\mu\nu}\eta_{\kappa\lambda}$$

and the average is

$$\langle |M|^2 \rangle = \frac{1}{4} \frac{g_e^4}{(p_4 - p_2)^4} \operatorname{Tr}\left[\gamma^\mu (\not{p}_1 + m_1 c) \bar{\gamma}^\kappa (\not{p}_3 + m_3 c)\right] \operatorname{Tr}\left[\gamma^\nu (\not{p}_2 + m_2 c) \bar{\gamma}^\lambda (\not{p}_4 + m_4 c)\right] \eta_{\mu\nu} \eta_{\kappa\lambda}$$

where the division by four comes from the two incoming particles with two spin states each.

3.4 Details for Calculations in Quantum Electrodynamics

The term $\gamma^\mu \not{P} = \gamma^\mu \gamma^\nu P_\nu$ in the spin average amplitude for the electron-muon scattering is $P_\nu \operatorname{Tr}[\gamma^\mu \gamma^\nu]$ because P_ν is a scalar in spin space and the trace is taken in spin space. The three equations

$$\begin{aligned} \operatorname{Tr}\left(\gamma^\mu \gamma^\nu \gamma^\kappa \gamma^\lambda\right) &= 4(\eta^{\mu\nu}\eta^{\kappa\lambda} - \eta^{\mu\kappa}\eta^{\nu\lambda} + \eta^{\mu\lambda}\eta^{\nu\kappa}) \\ \operatorname{Tr}(\gamma^\mu \gamma^\nu \gamma^\kappa) &= 0 \\ \operatorname{Tr}(\gamma^\mu \gamma^\nu) &= 4\eta^{\mu\nu} \end{aligned}$$

expressed in terms of the metric $\eta^{\mu\nu}$ help to evaluate the expression $|M|^2$ further.

One trace term in $|M|^2$ becomes

$$\begin{aligned} &\operatorname{Tr}\left[\gamma^\mu (\not{p}_a + m_a c) \bar{\gamma}^\kappa (\not{p}_b + m_b c)\right] \\ &= \operatorname{Tr}\left[\gamma^\mu \not{p}_a \gamma^\kappa \not{p}_b\right] + \operatorname{Tr}\left[\gamma^\mu \not{p}_a \gamma^\kappa m_b c\right] + \operatorname{Tr}\left[\gamma^\mu m_a c \gamma^\kappa \not{p}_b\right] + \operatorname{Tr}[\gamma^\mu m_a c \gamma^\kappa m_b c] \\ &= p_{\alpha a} p_{\beta b} \operatorname{Tr}\left(\gamma^\mu \gamma^\alpha \gamma^\kappa \gamma^\beta\right) + m_b c\, p_{\alpha a} \operatorname{Tr}(\gamma^\mu \gamma^\alpha \gamma^\kappa) + m_a c\, p_{\beta b} \operatorname{Tr}\left(\gamma^\mu \gamma^\kappa \gamma^\beta\right) + m_a m_b c^2 \operatorname{Tr}(\gamma^\mu \gamma^\kappa) \\ &= 4 p_{\alpha a} p_{\beta b} (\eta^{\mu\alpha}\eta^{\kappa\beta} - \eta^{\mu\kappa}\eta^{\alpha\beta} + \eta^{\mu\beta}\eta^{\alpha\kappa}) + 4 m_a m_b c^2 \eta^{\mu\kappa} = 4(p_a^\mu p_b^\kappa - \eta^{\mu\kappa} p_a \cdot p_b + p_a^\kappa p_b^\mu) + 4 m_a m_b c^2 \eta^{\mu\kappa} \end{aligned}$$

using $\bar{\gamma}^\mu = \gamma^0 \gamma^{\mu\dagger} \gamma^0 = \gamma^\mu$ and the definition of $\not{P} = \gamma^\mu P_\mu$ as well as $\operatorname{Tr}[A + B] = \operatorname{Tr} A + \operatorname{Tr} B$ and $\operatorname{Tr}[\alpha A] = \alpha \operatorname{Tr} A$ for matrices A and B and a scalar α.

The average for the electron-muon scattering is

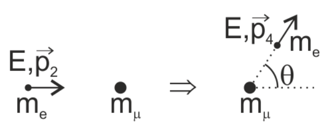
$$\begin{aligned} \langle |M|^2 \rangle &= \frac{1}{4} \frac{g_e^4}{(p_4 - p_2)^4} 16 \left[(p_1^\mu p_3^\kappa - \eta^{\mu\kappa} p_1 \cdot p_3 + p_1^\kappa p_3^\mu) + m_e^2 c^2 \eta^{\mu\kappa}\right] \\ &\qquad \left[(p_2^\nu p_4^\lambda - \eta^{\nu\lambda} p_2 \cdot p_4 + p_2^\lambda p_4^\nu) + m_\nu^2 c^2 \eta^{\nu\lambda}\right] \eta_{\mu\nu} \eta_{\kappa\lambda} \\ &= \frac{8 g_e^4}{(p_4 - p_2)^4} \left[(p_1 \cdot p_2)(p_3 \cdot p_4) + (p_1 \cdot p_4)(p_2 \cdot p_3) - (p_1 \cdot p_3) m_e^2 c^2 - (p_2 \cdot p_4) m_\mu^2 c^2 + 2 m_e^2 m_\mu^2 c^4\right] \end{aligned}$$

specified in terms of only 4-momenta and the masses of the electron and the muon. This is completely trivial in spin space.

With the approximation $m_\mu \gg m_e$ and the assumption $E_2 = E_e \ll m_\mu c^2$ the differential scattering cross section (2.5) for a two-body scattering in the center of momentum frame becomes

$$\langle \frac{d\sigma}{d\Omega} \rangle = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{S \langle |M|^2 \rangle}{(E_1 + E_2)^2} \frac{\left|\vec{P}_f\right|}{\left|\vec{P}_i\right|} \approx \left(\frac{\hbar}{8\pi m_\mu c^2}\right)^2 \langle |M|^2 \rangle$$

using $|\vec{P}_f| = |\vec{P}_i|$ and $E_1 + E_2 = m_\mu c^2 + E_2 \approx m_\mu c^2$ in the center of momentum frame which is the rest frame of the muon due to the huge difference in mass between electron and muon. (The collision of an electron with a muon is comparable with the collision of a tennis ball with the earth.)



One could plug the term above for $\langle |M|^2 \rangle$ into this expression, but one can simplify this expression using $p_1 = (m_\mu c, \vec{0})$, $p_2 = (\frac{E}{c}, \vec{p}_2)$, $p_3 = (m_\mu c, \vec{0})$, $p_4 = (\frac{E}{c}, \vec{p}_4)$ with $p = |\vec{p}_2| = |\vec{p}_4|$ since $m_\mu \gg m_e$. The dot-products become $p_1 \cdot p_3 = m_\mu^2 c^2$ and $p_2 \cdot p_4 = E^2/c^2 - \vec{p}_2 \cdot \vec{p}_4 = E^2/c^2 - p^2 \cos\theta$, the products of dot-products become $(p_1 \cdot p_2)(p_3 \cdot p_4) = (p_1 \cdot p_4)(p_2 \cdot p_3) = m_\mu^2 E^2$ and $(p_4 - p_2)^2 = 0 - (\vec{p}_4 - \vec{p}_2)^2 = -|\vec{p}_2|^2 - |\vec{p}_4|^2 + 2\vec{p}_2 \cdot \vec{p}_4$.

Using $\vec{p}_2 \cdot \vec{p}_4 = |\vec{p}_2||\vec{p}_4|\cos\theta = p^2\cos\theta = p^2 - p^2 + p^2\cos\theta = p^2 - 2p^2((1-\cos\theta)/2) = p^2 - 2p^2\sin^2(\theta/2)$ the two expressions $p_2 \cdot p_4$ and $(p_4 - p_2)^2$ become

$$p_2 \cdot p_4 = \left(\frac{E}{c}\right)^2 - \vec{p}_2 \cdot \vec{p}_4 = \left[\left(\frac{E}{c}\right)^2 - p^2\right] + 2p^2\sin^2\frac{\theta}{2} = m_e^2c^2 + 2p^2\sin^2\frac{\theta}{2}$$

$$(p_4 - p_2)^2 = -2p^2 + 2\vec{p}_2 \cdot \vec{p}_4 = -2p^2 + 2\left(p^2 - 2p^2\sin^2\frac{\theta}{2}\right) = -2p^2 + 2p^2 - 4p^2\sin^2\frac{\theta}{2} = -4p^2\sin^2\frac{\theta}{2}$$

and the differential scattering cross section becomes

$$\frac{d\sigma}{d\Omega} = \left(\frac{g_e^2\hbar}{8\pi p^2\sin^2\frac{\theta}{2}}\right)^2\left[(m_ec)^2 + p^2\cos^2\frac{\theta}{2}\right] \tag{3.4}$$

called the Mott formula or, in the non-relativistic limit with $p \ll (m_ec)^2$, gives

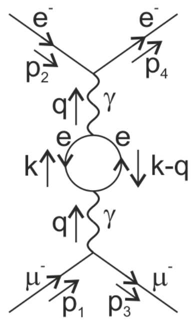
$$\frac{d\sigma}{d\Omega} = \left(\frac{g_e^2\hbar\, m_ec}{8\pi\, m_e^2v^2\,\sin^2\frac{\theta}{2}}\right)^2 = \left(\frac{e^2\, 4\pi\hbar\, m_ec}{8\pi\, \hbar c\, m_e^2v^2\,\sin^2\frac{\theta}{2}}\right)^2 = \left(\frac{e^2}{2m_ev^2\,\sin^2\frac{\theta}{2}}\right)^2$$

using $g_e = e\sqrt{4\pi/\hbar c} = \sqrt{4\pi\alpha}$ and therefore

$$\frac{d\sigma}{d\Omega} = \left(\frac{e^2}{2m_ev^2\sin^2\frac{\theta}{2}}\right)^2 \tag{3.5}$$

called Rutherford formula. One certainly expects the dependence on θ but no dependence on the angle perpendicular to the incoming beam.

3.5 Higher-Order Diagrams



The Mott formula is not yet the exact formula because only the lowest order diagram has been taken into account and some approximations have been made. The effects of virtual particle pairs start at fourth order with the largest contribution coming from a diagram with an internal loop of matter. (The internal momenta for the two photons could be labeled q_1 and q_2 but the delta-function would make them the same. Similarly the momenta q and $k-q$ come in at the lower internal vertex and momentum k goes out. Thus for internal momenta it is possible to label them such that calculations get easier.)

To evaluate diagrams with purely internal loops of matter, a new Feynman rule is needed: For internal loops of matter write down the ordered product of vertex factors and propagators (resulting in a matrix in spin space), then take the trace in spin space and multiply by -1.

Because no vertex factors with $u^{(s)}$ or $v^{(s)}$ are assigned to virtual fermions but only propagators, it is not important which of the two virtual electrons is called matter and which is called antimatter. The resulting quantity M with the contribution for the internal loop

$$M_{\text{loop}} = -\frac{ig_e^4}{q^4}[\bar{u}(3)\gamma^\mu u(1)]\left[\int \frac{d^4k}{(2\pi)^4}\frac{\mathrm{Tr}\left[\gamma_\mu(\not{k} + m_ec)\,\gamma_\nu(\not{k} - \not{q} + m_ec)\right]}{[k^2 - m_e^2c^2][(k-q)^2 - m_e^2c^2]}\right][\bar{u}(4)\gamma^\nu u(2)] \tag{3.6}$$

is the same.

The part in the large square brackets with the integral is actually divergent and will eventually lead to the topic of renormalization. The momentum exchanged at the top vertex is $q = p_2 - p_4$ but also the momentum $q = p_1 - p_3$ exchanged at the bottom vertex could have been used. Since the four external momenta are fixed, also q is fixed, but k is completely undetermined. This is the reason why the integral diverges. Renormalization is a systematic way to address these infinities as discussed below.

So far the fermions have been assumed to be leptons, but quarks and antiquarks behave very similarly to electrons, muons, tauons and their antiparticles in Quantum Electrodynamics. Everything stays the same except that the electric charge in the coupling constant g becomes $\frac{1}{3}e$ or $\frac{2}{3}e$ instead of simply e.

4 Application to Quantum Chromodynamics

4.1 Hierarchy of Interactions and Quarks with Electromagnetism

Looking at the interactivity of fermions, neutrinos are only affected by the weak interactions, charged leptons are affected by the electromagnetic and the weak interactions, and quarks are affected by the strong, the electromagnetic and the weak interactions. Charged leptons and quarks play a more important role in structure because it is hard to make a bound state with neutrinos. Charged leptons feel apart from the weak interactions also the electromagnetic force, and quarks in addition also the strong force.

These are only the direct interactions, but in a given process M all contributions must be included, and virtual states can bring unexpected additional interactions from other forces as shown in figure 4.1 for electron-muon scattering in (a). If one can draw the lowest order electromagnetic interaction with a photon γ as in (b), one can also draw the same diagram with a Z^0 boson for the weak interaction instead of the photon as in (c). Going to higher order diagrams one has to take into account all three interactions as in (d) where a virtual quark-antiquark pair is created and annihilated. Since quarks can exchange gluons also such interactions between virtual quarks as in diagram (e) are possible although they have much less influence than the lower order diagrams (b) and (c).

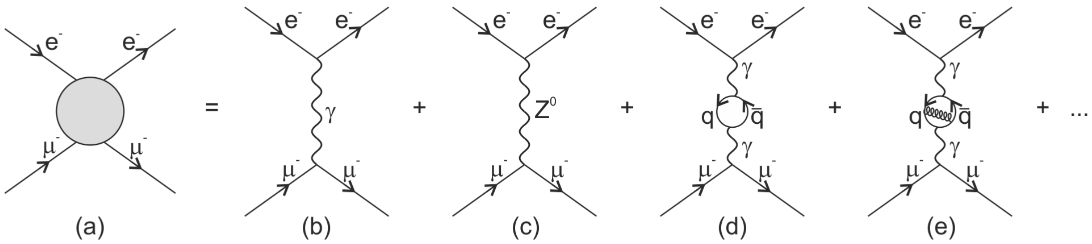


Figure 4.1: Additional interactions from virtual states

In the following only the lowest order diagrams of Quantum Chromodynamics are considered. However, since quarks are charged particles also electromagnetic interactions are important. Quarks in electromagnetism can be handled the same way as the charged leptons. The only difference is that they do not have integer electric charge such that $\pm\frac{1}{3}e$ or $\pm\frac{2}{3}e$ replaces e in the coupling constant g_e.

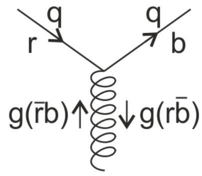
4.2 Feynman Diagrams for Quantum Chromodynamics

The Lagrangian for Quantum Chromodynamics is

$$\begin{aligned}\mathcal{L} = &\sum_{i=1}^{6}\left[\hbar c\bar{\Psi}_i\gamma^\mu D_\mu\Psi_i + m_i c^2\bar{\Psi}_i\Psi_i\right] + \frac{1}{16\pi}\left(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a\right)\left(\partial^\mu A^{\nu a} - \partial^\nu A^{\mu a}\right)\\ &- \frac{g}{16\pi}f^{ade}A_\mu^d A_\nu^e\left(\partial^\mu A^{\nu a} - \partial^\nu A^{\mu a}\right) - \frac{g}{16\pi}f^{abc}A^{\mu b}A^{\nu c}\left(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a\right)\\ &+ \frac{g^2}{16\pi}f^{abc}f^{ade}A_\mu^b A_\nu^c A^{\mu d}A^{\nu e}\end{aligned}$$

where $D_\mu\Psi = \partial_\mu + ig\lambda\cdot A_\mu\Psi$ with the eight generators λ of SU(3) and the sum goes over all six flavors of quarks. The Dirac Lagrangian handles quarks and antiquarks such that the antiquarks do not have to be listed separately. Each Ψ_i is a vector in color space with the three components Ψ_{ri}, Ψ_{bi}, Ψ_{g_i}, and the three colors (r, b, g) play the role of charges and interactions via eight gluons. This and the non-abelian character of SU(3) makes it already more complicated than Quantum Electrodynamics with one type of charge instead of three and one photon instead of eight gluons. Another complication comes from the fact that Quantum Chromodynamics is defined in terms of quarks, but one can only observe and experiment the bound states which are hadrons.

The first line of the Lagrangian looks similar to the case for Quantum Electrodynamics, but the second and third line of the Lagrangian come in addition representing the gluon-gluon interactions. Therefore there will be Feynman diagrams with pure gluon vertices. The two terms in the second line lead to vertices with three gluons and the term in the third line results in vertices with four gluons.



The vertex showing the interaction of a quark with a gluon looks very similar to the vertex showing an electron interacting with a photon. However there are differences. The photon is not electrically charged, but the gluons have color. The electron comes in and goes out with the same electrical charge, but the quark changes the color charge. Therefore one also has to label the color charges. An incoming red quark, for example, may become through the interaction blue. There are two ways to interpret the gluon. One way is that the gluon moves into the vertex carrying $\bar{r}$ to cancel the red but also carrying b to set blue, but another way is that the gluon moves out of the vertex carrying r and canceling b with $\bar{b}$. Thus one draws an arrow, and it does not matter in which direction it points because it only indicates which of the two possibilities are chosen.

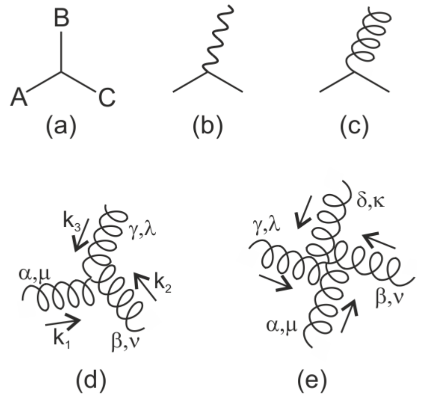
One can think of gluons as something that carries one color and one anticolor. Also mesons come in quark-antiquark pairs and therefore carry color and anticolor. There are eight octet states and one singlet state

$$
\begin{aligned}
|1\rangle &= \frac{1}{\sqrt{2}}(r\bar{b}+b\bar{r}) & |5\rangle &= -\frac{i}{\sqrt{2}}(r\bar{g}-g\bar{r}) & |9\rangle &= \frac{1}{\sqrt{3}}(r\bar{r}+b\bar{b}+g\bar{g}) \\
|2\rangle &= -\frac{i}{\sqrt{2}}(r\bar{b}-b\bar{r}) & |6\rangle &= \frac{1}{\sqrt{2}}(b\bar{g}+g\bar{b}) & & \\
|3\rangle &= \frac{1}{\sqrt{2}}(r\bar{r}-b\bar{b}) & |7\rangle &= -\frac{i}{\sqrt{2}}(b\bar{g}-g\bar{b}) & & \\
|4\rangle &= \frac{1}{\sqrt{2}}(r\bar{g}+g\bar{r}) & |8\rangle &= \frac{1}{\sqrt{6}}(r\bar{r}+b\bar{b}-2g\bar{g}) & &
\end{aligned}
\tag{4.1}
$$

where $|1\rangle$ to $|8\rangle$ are the octet states and $|9\rangle$ is the singlet state. The singlet state is colorless and invariant under SU(3), and the eight octet states get rotated into each other under SU(3) where the states $|3\rangle$ and $|8\rangle$ are in a non-trivial way colorless. It would be much nicer to work in the basis of the nine states $r\bar{r}$, $r\bar{b}$ and so on, but this would make it difficult to see the transformation properties and to figure out which is the singlet state. However the simpler states are often used in Feynman diagrams and can easily be expressed as linear combinations of the states in the basis of (4.1).

In terms of the physics, all mesons are color singlets, and all gluons are color octet states. (Note that the λ in the term $\lambda \cdot A_\mu$ of the definition of the covariant derivative secretly carries two color indices because it is a 3×3 matrix in color space and an index a specifying which generator it is.) The state $|9\rangle$ does not exist as a gluon. If it would, the gauge group for the strong force would not be SU(3) but U(3). If there would be a U(3) gauge theory the $|9\rangle$ boson would exactly act like the photon in the sense that a quark could come in and go out without effect on the color. It has never been detected though and is therefore experimentally ruled out.

Comparison of the theories so far:



	External State Labels	Internal Propagators	Vertex Factors	
ABC	none	$\frac{i}{q^2-m^2c^2}$	$-ig$	(a)
QED	u, $\bar{u}$, v, $\bar{v}$ "matter"	$\frac{i(\not{q}+mc)}{q^2-m^2c^2}$	$ig_e\gamma^\mu$	(b)
	ϵ_μ, ϵ^*_μ "photons"	$-\frac{i\eta_{\mu\nu}}{q^2}$		
QCD	uc, $\bar{u}c^\dagger$, vc, $\bar{v}c^\dagger$ "matter"	$\frac{i(\not{q}+mc)}{q^2-m^2c^2}$	$-\frac{ig_s}{2}\lambda^\alpha\gamma^\mu$	(c)
	$\epsilon_\mu a^\alpha$, $\epsilon^*_\mu a^{\alpha *}$ "gluons"	$-\frac{i\eta_{\mu\nu}\delta^{\alpha\beta}}{q^2}$	$-g_s f^{\alpha\beta\gamma}\left[\eta_{\mu\nu}(k_1-k_2)_\lambda+\eta_{\nu\lambda}(k_2-k_3)_\mu+\eta_{\lambda\mu}(k_3-k_1)_\nu\right]$	(d)
			$-ig_s^2\left[f^{\alpha\beta\rho}f^{\gamma\delta\rho}(\eta_{\mu\lambda}\eta_{\nu\kappa}-\eta_{\mu\kappa}\eta_{\nu\lambda})+ f^{\alpha\delta\rho}f^{\beta\gamma\rho}(\eta_{\mu\nu}\eta_{\lambda\kappa}-\eta_{\mu\lambda}\eta_{\nu\kappa})+f^{\alpha\gamma\rho}f^{\delta\beta\rho}(\eta_{\mu\kappa}\eta_{\nu\lambda}-\eta_{\mu\nu}\eta_{\lambda\kappa})\right]$	(e)

where the ABC theory is a toy theory but no gauge theory, QED is Quantum Electrodynamics, and QCD is Quantum Chromodynamics.

There are three colors c appearing in the external state labels under QCD together with the solutions of the Dirac equation (3.2). The charge space for Quantum Chromodynamics is three-dimensional with

$$c_1 = \begin{pmatrix}1\\0\\0\end{pmatrix} = \text{red} \qquad c_2 = \begin{pmatrix}0\\1\\0\end{pmatrix} = \text{blue} \qquad c_3 = \begin{pmatrix}0\\0\\1\end{pmatrix} = \text{green}$$

such that an arbitrary state can be written as $c = a_i c_i$ with complex coefficients a_i. Hence for the color c the anticolor is $c^\dagger = c^{T*}$. The three c_i build a basis in the H^3 space of color states.

There are eight gluons a^α appearing in the external state labels under QCD with

$$a^1 = \begin{pmatrix}1\\0\\0\\0\\0\\0\\0\\0\end{pmatrix} \quad a^2 = \begin{pmatrix}0\\1\\0\\0\\0\\0\\0\\0\end{pmatrix} \quad a^3 = \begin{pmatrix}0\\0\\1\\0\\0\\0\\0\\0\end{pmatrix} \quad a^4 = \begin{pmatrix}0\\0\\0\\1\\0\\0\\0\\0\end{pmatrix} \quad a^5 = \begin{pmatrix}0\\0\\0\\0\\1\\0\\0\\0\end{pmatrix} \quad a^6 = \begin{pmatrix}0\\0\\0\\0\\0\\1\\0\\0\end{pmatrix} \quad a^7 = \begin{pmatrix}0\\0\\0\\0\\0\\0\\1\\0\end{pmatrix} \quad a^8 = \begin{pmatrix}0\\0\\0\\0\\0\\0\\0\\1\end{pmatrix}$$

which build a basis in the H^8 space of gluon states. They correspond to the eight generators of SU(3). The ninth gluon corresponding to the singlet state $|9\rangle$ in (4.1) has not been observed. Of course, also the other eight gluons have not been seen but for those one can use the confinement argument discussed below stating that only color singlets can be observed.

For Quantum Electrodynamics the four 4×4 γ^μ_{ab} matrices linked the spin space to spacetime. In a similar way the eight 3×3 λ^μ_{ij} matrices of Quantum Chromodynamics link the color space to the gluon space. (They are the generators of SU(3), each is a 3×3 matrix, and there are eight of them.) The indices i and j are usually not written, but one should keep in mind that they are matrices based on matrix multiplication where the order is important. The colors c and $c^\dagger$ are used together with the λ matrices in the same way as the γ^μ are used in spinor sandwiches, and the result will be a scalar.

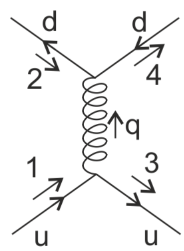
The 512 structure constants $f^{\alpha\beta\gamma}$ come from the fact that SU(3) is not abelian. They appear in the Lie algebra in $[\lambda^\alpha, \lambda^\beta] = 2if^{\alpha\beta\gamma}$.

A big technical complication arises for diagrams involving internal loops. In this case one has to be very careful not to count gauge equivalent and therefore physically indistinct configurations more than once. To get the counting right without losing gauge invariance, one introduces Faddeev–Popov ghosts which are additional non-physical fields whose sole purpose is to cancel the non-physical gauge equivalent fluctuations of the physical fields.

4.3 Mesons as a Pair of a Quark and an Antiquark

One question is why one only finds color singlet-bound states for matter, or, in other words, why are quarks always bound up in hadrons. One type of the hadrons are the mesons which consist of a quark q_1 and an antiquark $\bar{q}_2$ which do not need to be of the same flavor. Because quarks are fractionally charged a meson may or may not be electrically neutral.

The idea is to see if the strong force is attractive leading to bound states or repulsive leading to no bound states. One expects attraction for colorless singlet combinations and repulsion for colorful octet combinations. Treating this as a scattering event $u+\bar{d} \to u + \bar{d}$ should answer the question how the up quark and the down antiquark interact by exchanging a gluon. What one finds out is that when two quarks are bound in a meson or three quarks are bound in a baryon they actually behave like free particles in energy scales associated with the separation between them. That has a lot to do with how the strong force varies with energy scales. Thus treating the situation as a scattering event is the right approach, and one can just apply the Feynman rules.



One gets after canceling $(2\pi)^4\delta^4(p_{\text{tot in}} - p_{\text{tot out}})$ and multiplication by i

$$M = i\,\bar{u}(3)c^\dagger(3)\big(-i\frac{g_s}{2}\lambda^\alpha\gamma^\mu\big)u(1)c(1)\,\big[-\frac{i\eta_{\mu\nu}\delta^{\alpha\beta}}{q^2}\big]\bar{v}(2)c^\dagger(2)\big(-i\frac{g_s}{2}\lambda^\beta\gamma^\nu\big)v(4)c(4)$$
$$= -\frac{g_s^2\eta_{\mu\nu}}{q^2}\,\bar{u}(3)\gamma^\mu u(1)\bar{v}(2)\gamma^\nu v(4)\,\big[\big(c^\dagger(3)\lambda^\alpha c(1)\big)\big(c^\dagger(2)\lambda^\alpha c(4)\big)\big]\cdot\frac{1}{4}$$

because λ^α and λ^β are not matrices in spin space. The two factors in the square bracket of the second line are color sandwiches giving numbers. Thus one can separate the spinor sandwiches from the color sandwiches. The part with the spinor sandwiches is exactly what one gets from Quantum Electrodynamics with $g_e \to g_s$ from, for example, $e^- + \mu^+ \to e^- + \mu^+$.

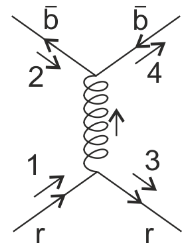
The color factor in square brackets (with the two color sandwiches) which in the end is a number contains all the differences between Quantum Electrodynamics and Quantum Chromodynamics. Therefore $|M|^2$ can be calculated and averaged the same way as in Quantum Electrodynamics, and one can concentrate on the evaluation of the color factor f. In a sense one can model the scattering of an electron e^- and an antimuon μ^+ with a potential $V(r)$ and calculate $\langle\Psi_{\text{final}}|V(r)|\Psi_{\text{initial}}\rangle$. Attraction versus repulsion is determined by the sign of $V(r)$. Thus in the case of a negatively charged electron and a positively charged antimuon it is attractive. Therefore in the case of a quark and an antiquark the color factor determines with $f > 0$ attraction and with $f < 0$ repulsion. It is not clear whether a red and an antiblue quark, for example, give an attractive or a repulsive interaction. It is the color factor f which shows what it is.

The nine states in (4.1) turned out to be $3 \otimes \bar{3} = 8 \oplus 1$ with eight colorful octet states and one colorless singlet state. In the following, the color factor f is investigated choosing either one of the octet states where one expects $f < 0$ or the singlet state where one expects $f > 0$ because mesons are only allowed to be in a singlet state. The quantities λ^α

$$\lambda^1 = \begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix} \quad \lambda^2 = \begin{pmatrix}0&-i&0\\i&0&0\\0&0&0\end{pmatrix} \quad \lambda^3 = \begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix} \quad \lambda^4 = \begin{pmatrix}0&0&1\\0&0&0\\1&0&0\end{pmatrix}$$
$$\lambda^5 = \begin{pmatrix}0&0&-i\\0&0&0\\i&0&0\end{pmatrix} \quad \lambda^6 = \begin{pmatrix}0&0&0\\0&0&1\\0&1&0\end{pmatrix} \quad \lambda^7 = \begin{pmatrix}0&0&0\\0&0&-i\\0&i&0\end{pmatrix} \quad \lambda^8 = \frac{1}{\sqrt{3}}\begin{pmatrix}1&0&0\\0&1&0\\0&0&-2\end{pmatrix} \tag{4.2}$$

are the matrices in color space and represent the eight generators of SU(3).

If the quark-antiquark pair $u\bar{d}$ were in a colorful octet state such as $r\bar{b} = \frac{1}{\sqrt{2}}(|1\rangle + i\,|2\rangle)$ then



$$c(1) = \begin{pmatrix}1\\0\\0\end{pmatrix} \quad c(2) = \begin{pmatrix}0\\1\\0\end{pmatrix} \quad c(3) = \begin{pmatrix}1\\0\\0\end{pmatrix} \quad c(4) = \begin{pmatrix}0\\1\\0\end{pmatrix}$$

are the colors needed for the color factor f. Gluons can change color but they do not have to change color at a vertex. In this case it does not because of color conservation. The color factor becomes

$$f = \frac{1}{4}\sum_{\alpha=1}^{8}\left((1,0,0)\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}\right)\left((0,1,0)\lambda^\alpha\begin{pmatrix}0\\1\\0\end{pmatrix}\right) = \frac{1}{4}\left(0+0-1+0+0+0+0+\frac{1}{3}\right) = -\frac{1}{6}$$

and is negative such that this implies repulsion between a quark and an antiquark in this color octet state. (Note that the first factor in f grabs λ^α_{11} and the second factor λ^α_{22} such that only λ^3 and λ^8 contribute.) Surprisingly any other octet state gives the same number $-\frac{1}{6}$. This means that if there are two quarks in a color octet state they cannot be in a bound state.

If one uses the color singlet state $|9\rangle = \frac{1}{\sqrt{3}}(r\bar{r} + b\bar{b} + g\bar{g})$ instead which is notably not $(r+b+g)(\bar{r}+\bar{b}+\bar{g})$ then the calculations become a bit more difficult because $|9\rangle$ is not simply a product of some colors. Using

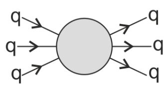
$$c(1)c(2) = \frac{1}{\sqrt{3}}\left[\begin{pmatrix}1\\0\\0\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix} + \begin{pmatrix}0\\1\\0\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix} + \begin{pmatrix}0\\0\\1\end{pmatrix}\begin{pmatrix}0\\0\\1\end{pmatrix}\right] = c(3)c(4)$$

the color factor becomes

$$
\begin{aligned}
f &= \frac{1}{4}\sum_{\alpha=1}^{8}\left(c^\dagger(3)\lambda^\alpha c(1)\right)\left(c^\dagger(2)\lambda^\alpha c(4)\right) \\
&= \frac{1}{4}\frac{1}{\sqrt{3}}\sum_{\alpha=1}^{8} c^\dagger(3)\lambda^\alpha \left[\begin{pmatrix}1\\0\\0\end{pmatrix}(1,0,0)+\begin{pmatrix}0\\1\\0\end{pmatrix}(0,1,0)+\begin{pmatrix}0\\0\\1\end{pmatrix}(0,0,1)\right]\lambda^\alpha c(4) \\
&= \frac{1}{4}\frac{1}{\sqrt{3}}\frac{1}{\sqrt{3}}\sum_{\alpha=1}^{8}\left[(1,0,0)\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}(1,0,0)\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}+(0,1,0)\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}(1,0,0)\lambda^\alpha\begin{pmatrix}0\\1\\0\end{pmatrix}\right. \\
&\qquad\qquad \left.+(0,0,1)\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}(1,0,0)\lambda^\alpha\begin{pmatrix}0\\0\\1\end{pmatrix}+(1,0,0)\lambda^\alpha\begin{pmatrix}0\\1\\0\end{pmatrix}(0,1,0)\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}+\ldots\right] \\
&= \frac{1}{12}\sum_{\alpha=1}^{8}\lambda^\alpha_{ij}\lambda^\alpha_{ji} = \frac{1}{12}\sum_{\alpha=1}^{8}\mathrm{Tr}(\lambda^\alpha\lambda^\alpha) = \frac{4}{3}
\end{aligned}
$$

showing that the interaction between a quark and an antiquark in the singlet configuration is attractive. (Note that the flavor of the quark and antiquark does not show up in the calculation and only the colors were relevant.)

4.4 Baryons as a Bound State of Three Quarks



Baryons consist of three quarks q_1, q_2, q_3 (and antibaryons of three antiquarks). To show that the force between the three quarks is attractive when they form a color singlet and repulsive otherwise is considerably more tedious. One might think that a scattering process with three incoming and three outgoing quarks is needed to evaluate, but this is complicated and a bit misleading. The problem is that one might get some overall attraction but with a repulsion between two of the quarks. To analyze this it is easier to first consider pairs of quarks similar to the quark-antiquark scattering but this time as quark-quark scattering. If the three quarks in a baryon attract each other pairwise, then one can assume that all three attract each other although it is not completely impossible that there is a repulsive force between all three which is stronger than the pairwise attraction. One gets

$$
M = -\frac{g_s^2}{4}\frac{1}{q^2}\left[\bar{u}(3)\gamma^\mu u(1)\right]\left[\bar{u}(4)\gamma^\mu u(2)\right]\left(c^\dagger(3)\lambda^\alpha c(1)\right)\left(c^\dagger(4)\lambda^\alpha c(2)\right)
$$

which is the same as for the repulsive electron-muon scattering except for the color factor. This time the color factor has to be negative for attraction and positive for repulsion.

To evaluate the color factor combinations of color assignments for two quarks must be considered. For a quark and an antiquark the colors $c_1\bar{c}_2$ gave $3\otimes\bar{3} = 8\oplus 1$ with the octets and the singlet. For two quarks the colors c_1c_2 give $3\otimes 3 = 3\oplus 6$ with a triplet and a sextet

$$
\begin{aligned}
\text{triplet} : &\left\{\frac{1}{\sqrt{2}}(rb-br), \frac{1}{\sqrt{2}}(gr-rg), \frac{1}{\sqrt{2}}(bg-gb)\right\} \\
\text{sextet} : &\left\{rr, bb, gg, \frac{1}{\sqrt{2}}(rb+br), \frac{1}{\sqrt{2}}(gr+rg), \frac{1}{\sqrt{2}}(gb+bg)\right\}
\end{aligned}
$$

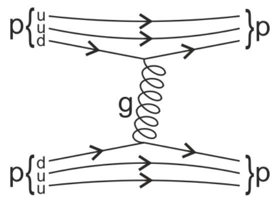
where the triplet states are antisymmetric under the exchange $c_1 \leftrightarrow c_2$ and the sextet states are symmetric under the exchange $c_1 \leftrightarrow c_2$. The color factors for any of the triples is $f = -\frac{2}{3}$ showing attraction and for any of the sextets $f = \frac{1}{3}$ showing repulsion.

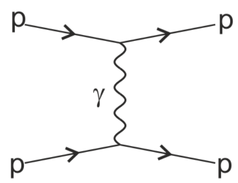
This is only for two quarks but there are three quarks in a baryon. Taking two quarks in a baryon then they have to be in a triplet state. However this does not specify the color state of all three quarks. For three quarks one gets $3\otimes 3\otimes 3 = 10\oplus 8\oplus 8\oplus 1$ where the 10 contains the totally symmetric decuplet states with rrr or normalized sum of all six permutations of rbg and the 1 is the totally antisymmetric singlet state counting even permutations of rbg positively and odd permutations negatively. The two 8 contain the states mixing symmetric and antisymmetric parts where a triplet state of two quarks, for example, is multiplied by a color. Only in the totally antisymmetric singlet configuration is the force between any two pairs of quarks in the baryon attractive. Therefore to reliably predict bound states one needs the quarks in a hadron to form a color singlet.

These are only suggested calculations and no rigorous proofs that there cannot be a bound state with a repulsive force between two of the quarks and so on. On the other hand, the SU(3) symmetry only allows colorful octet gluons. There are two experimental observations that one should consider here. One has never seen free quarks, and one has contrary to photons never seen free gluons. The color confinement hypothesis says that any free or long lived state in Quantum Chromodynamics must be in a color singlet configuration. This rules out gluons and single quarks but allows mesons and baryons[1]. Proving this hypothesis is one of the great outstanding problems in Particle Physics. For reasons discussed when studying renormalization it has a lot to do with the breakdown of perturbation theory in certain situations such that the Feynman approach is useless because for large distances the coupling constant of the strong force becomes greater than one. Exact methods like lattice gauge theory, supersymmetry and dualities might eventually shed some light. Here it is assumed that the hypothesis is correct.

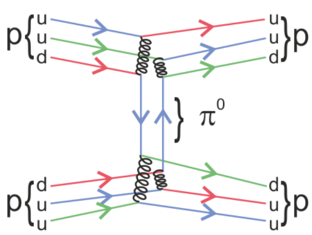
4.5 The Weakness of the Strong Interaction

This leads to the question why the strong force is swamped by electromagnetism at large distance if the strong force is so strong. At large distances the effects of electromagnetism are observable but those of the strong force not. One cannot calculate M when a quark and an antiquark are separated by a fair distance because $g > 1$ such that the higher order vertices count more and the infinity of diagrams cannot be ignored after a finite number of diagrams. When the quark and the antiquark are really close together Quantum Chromodynamics gets weak which is called asymptotic freedom. However, when they are close together they do not even feel each other. It is exactly opposite to electromagnetism.





Thus assuming two protons separated by a reasonable distance and the interaction between them one can model the interaction by a gluon exchange. Both protons p have two up quarks u and one down quark d, and the two down quarks may exchange the gluon. One can compare this proton-proton interaction and the exchange of a gluon with the electromagnetic proton-proton interaction and the exchange of a photon. The gluon exchange is $O(\alpha_S^2)$ and the photon exchange is $O(\alpha_E^2)$. This together with the fact that the strong coupling α_S is larger than the weak coupling α_W which is larger than the electromagnetic coupling α_E or written as the inequality $\alpha_S > \alpha_W > \alpha_E$ leads to the conclusion that the gluon exchange is more important than the photon exchange, but one can measure the electromagnetic force between two photons separated by one meter and not the strong force. The electromagnetic force is therefore much stronger such that the question arises whether the electromagnetic amplitude is much larger than the one of the strong interaction.



When two protons are separated by such a distance, the colorful gluon has to be long-lived, but the argument of confinement is that this is not possible. The two protons would have to exchange a singlet. One possibility is the exchange of a pion π^0 which as a meson is a color singlet and consists of a down quark and an down antiquark. This is $O(\alpha_S^8)$ compared with the Quantum Electrodynamics process which is $O(\alpha_E^2)$. So this will usually be swamped by the electromagnetic version. This picture is simplified, in part because each proton p is really in $1/\sqrt{6}(rgb - rbg - grb + +brg + gbr - bgr)$ while the π^0 is in $1/\sqrt{3}(r\bar{r} + b\bar{b} + g\bar{g})$. This process is not the exchange of a pion π^0 but pure Quantum Chromodynamics.

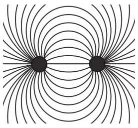
If one wants to see how this looks for a proton-neutron process, one has to replace an incoming and an outgoing up quark with a down quark at the bottom of the diagram. This does not change the diagram

[1]Photons do not interact with each other, but gluons do, and therefore bound states of gluons called glueballs could be possible although they have so far not been observed with certainty.

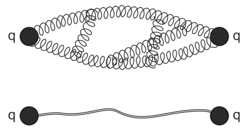
otherwise but there is no electromagnetic competition anymore because the neutron is electrically neutral. That is why one gets more stable nuclei by putting more and more neutrons into the nucleus. Adding more protons produce more electromagnetic repulsion, and adding more neutrons neutralizes that to a certain extent.

The reason why the strong interactions are less important than the electromagnetic interactions at large distances is because the diagrams are more complicated. Simple gluon exchange similar to photon exchange does not exist.

Before this was understood and Quantum Chromodynamics was there, it was clear that the protons are made out of something because through scattering experiments it was known that they start wiggling and that there are three things inside of them. However it was not possible to break them apart and similarly mesons could not be broken apart. Thus, the idea was to separate two quarks and see what happens compared to the electromagnetic field lines.



The early strategy to explain the absence of free quarks was based on the idea that the force between separated quarks gets confined to a tube connecting them and collected in a bundle, and that the field lines do not spread out as the electromagnetic field lines do in Quantum Electrodynamics. These color flux tubes may very well occur in Quantum Chromodynamics in part due to the gluon-gluon attraction. It is known that the gluons attract each other and may very well pull these field lines together.



The explanation why one cannot separate quarks is because the strong interaction when one tries to pull the two quarks apart is collapsing to such a tube between them and a tube has tension. If one pulls harder one raises the energy. Contrary to the electromagnetic field lines which get weaker the larger the distance is between the two charged particles, the strong interaction gets stronger the larger the distance. That is why one cannot pull them apart.

Studying the dynamics of these flux tubes in Quantum Chromodynamics in the scale of $\sim 10^{-15}$ m to explain these tubes led to problems like tachyons, spin-2 excitations and so on. When Quantum Chromodynamics was invented, one had to think about these tubes. Eventually one tried to shrink them to $\sim 10^{-35}$ m such that the spin-2 state could be the graviton, and by adding supersymmetry also the tachyons disappeared. This was the beginning of String Theory.

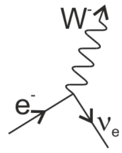
5 Application to the Weak Interaction

5.1 The Weak Interaction Compared to the Other Interactions

Electromagnetism is known for a long time and has been studied for over hundred years while it took much longer to get an understanding of the strong and weak interactions. It certainly is much easier than the strong and weak interactions. For Quantum Chromodynamics the difficulty is that quarks and gluons are bound, and the weak interactions are even more complicated because of the horrible Lagrangian with spontaneously broken symmetry, but they are very important because of the true particle decays. Despite the name weak interactions are not that weak because $\alpha_S > \alpha_W > \alpha_E$ at low energies, but one finds that the weak interaction related amplitudes are suppressed because the $W^{\pm}$ and the Z^0 are so massive. If one has propagators internal to a diagram, the mass suppresses this propagator strongly.

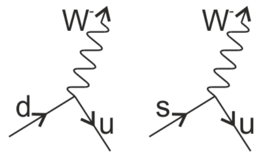
The weak interactions have an ugly side. It will be described following a historical development because there is a beautiful lesson in this.

5.2 Discovery of the Charm Quark



The two charged bosons $W^{\pm}$ change lepton flavor within single generations. The three lepton generations are e with ν_e, μ with ν_μ, τ with ν_τ each building a doublet. When the incoming particle is an electron e^-, the outgoing lepton must be an electron-neutrino ν_e, and because of charge conservation a W^- is also leaving the vertex.

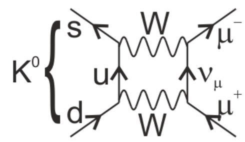
The situation for quarks is a bit different. The three generations consist of the doublets u and d, c and s, t and b. The first quark in a doublet carries the charge $\frac{2}{3}$, and the second quark in the doublet the charge $-\frac{1}{3}$. For increasing mass the order is u, d, s, c, b, t, and this is therefore also the order of discovery because more mass needs more energy.



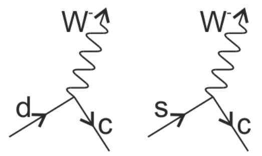
In 1963 only the u, d, s quarks were known, but to describe the results of experiments Cabibbo realized that both d and s should be related to u and quantified this with

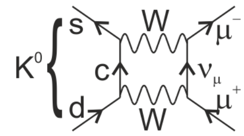
$$-\frac{ig_W}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)\cos\theta_c \qquad\qquad -\frac{ig_W}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)\sin\theta_c$$

where the left is the dominant contribution. In fact the Cabibbo angle θ_c is $\theta_c = 13.5°$. Note that the diagram with the s coming in and the u going out is changing generation. (The diagrams are not the full interaction but only an inner part.) With this angle Cabibbo could construct an agreement with the experimental facts.



When allowing cross-generation weak interactions one has to consider new processes. There was a particle called K^0, the neutral K-meson, which consists of a strange antiquark and a down quark. Now with the proposal of Cabibbo a process $K^0 + \bar{s} + d \to \mu^+ + \mu^-$ was possible where the down quark becomes an up quark and then the up quark becomes a strange quark mediated by two W bosons together with a leptonic process. If the cross-generation diagram with $\sin\theta_c$ is allowed also this process must be possible, and M comes with a factor $\cos\theta_c \sin\theta_c$. The problem is that one can calculate the amplitude associated with this and use it as a first-order approximation to figure out the lifetime of the K^0, but this does not match the observed lifetime of the K^0. However, it lives much longer. The question is how to solve this puzzle.





In 1970 Glashow, Iliopoulos and Maiani proposed a new quark called charm giving two complete generations of quarks. The corresponding diagram results in an M that comes with a canceling factor $-\cos\theta_c \sin\theta_c$. Because of the different mass of the up and the charm quark, they do not cancel completely. Since the K^0 does decay but with a lower decay rate, this is fine. The two diagrams corresponding to the one Cabibbo suggested give

$$-\frac{ig_W}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)(-\sin\theta_c) \qquad\qquad -\frac{ig_W}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)\cos\theta_c$$

as contributions. This is called the GIM-mechanism. The charm quark has been discovered in 1974.

Defining $d' = d\cos\theta_c + s\sin\theta_c$ and $s' = -d\sin\theta_c + s\cos\theta_c$ then the $W^\pm$ rotate within doublets of the form

$$\begin{pmatrix} u \\ d' \end{pmatrix} \text{ and } \begin{pmatrix} c \\ s' \end{pmatrix}$$

and the question is why is that.

5.3 Electrically Charged Weak Interactions

The transformation from d, s to d', s' is clearly a rotation of some kind, but moreover it takes for example a purely d' state and expresses it as a superposition of d and s. One might think that one can forget d and s and just use d' and s' instead, or one might ask where else in physics one does encounter this type of behavior where a single state is being turned into a superposition state. One possibility that might come up is measuring non-commuting observables in Quantum Mechanics. Taking for example X and P_x with $[X, P_x] \neq 0$ then one is in a superposition of P_x if one is in an eigenstate of X or vice versa.

The root of the problem is that quarks propagate as eigenstates of the free Dirac Hamiltonian $\hat{H}_D$, but they decay due to interactions mediated by $W^\pm$ and associated with weak vertex operators $\hat{V}_{W^\pm}$ where $[\hat{H}_D, \hat{V}_{W^\pm}] \neq 0$ such that eigenstates of one operator are not simultaneously eigenstates of the other.

After the charm quark has been proposed but was not yet discovered, Kobayashi and Maskawa introduced the third generation of quarks with the top and bottom quark and extended the Cabibbo rotation to a 3×3 transformation. The corresponding complex matrix

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} v_{ud} & v_{us} & v_{ub} \\ v_{cd} & v_{cs} & v_{cb} \\ v_{td} & v_{ts} & v_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{5.1}$$

is called the CKM-matrix because of Cabibbo, Kobayashi, Maskawa. To leading order this matrix is the identity I because all the off-diagonal terms are small compared to the diagonal terms.

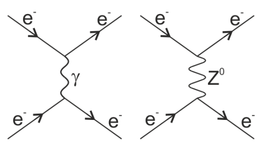
They had actually a good reason for doing so. They really wanted a truly complex quark rotation matrix, but for a 2×2 matrix it could be made real by redefining quark phases. With a 3×3 matrix though it could be complex and not be rendered real. The complex element in this matrix is the root of the CP-violation in the Standard Model that had been measured but was not understood in 1964. The motivation of the work of Kobayashi and Maskawa was the baryonic asymmetry of the universe. They hypothesized a third generation of quarks, and eventually as of 1995 all of the proposed three generations of quarks have been produced in collider experiments.

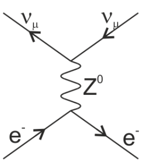
5.4 Electrically Neutral Weak Interaction

The CKM-matrix is first of all connected to quarks, but secondly it is restricted to charged weak interactions with the $W^{\pm}$ bosons only. The electrically neutral interaction with the Z^0 boson has also an interesting structure.

The Z^0 was first theoretized in 1958 by Bludman and put on solid theoretical ground by Glashow in 1961 when he unified electromagnetism and the weak interactions into a single electroweak theory. In 1967 Weinberg and Salaam finished the story by exploring why the interactions appear so different today. Altogether this forms the Glashow-Weinberg-Salaam electroweak theory in which the Higgs mechanism plays a key role. It took until 1973 to get experimental confirmation of the Z^0 boson.

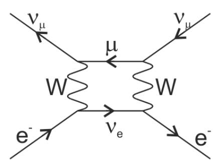
The reason why it took so long is that the photon γ can be replaced by a Z^0 in any Feynman diagram of Quantum Electrodynamics, but $M_E \gg M_W$ so that the weak process is always overshadowed. (Note that γ and Z^0 are both their own antiparticle.) The coupling in the weak vertex is larger, but the mass of the Z^0 is huge, and the probability for the process with γ is so much larger than the one for Z^0 that it is very difficult to observe the weak interaction.





However the reverse is not true and there are some weak diagrams with Z^0 that cannot be replaced with γ. Thus one has to find a process with Z^0 where the weak gauge boson cannot be exchanged with a photon, and such a process involves some neutral particles. A neutrino-electron scattering, for example, would satisfy the required condition but it is a process which is not easy to detect. In 1973 it had been observed at CERN though.

One might argue that the process seen at CERN really was the exchange of charged W bosons in a more complex diagram involving four vertices instead of two, but this is fourth order compared to Z^0 and would have come with a much smaller amplitude. Thus the observation at CERN showed a Z^0 and it was a bit tricky because it involved neutrinos.



Because nothing in the weak interactions is easy, this is not the end of the story. It turned out that the Z^0 boson has an additional layer of complications. A convenient property of it is that it does not change flavor and the CKM mechanism is therefore not involved. The usual vertex factor however is

$$-\frac{ig_W}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)$$

modulo CKM factors or quark composite corrections, and the Z^0 vertex factor takes a different form for each flavor f

$$-\frac{ig_Z}{2}\gamma^\mu(c_V^f - c_A^f\gamma^5)$$

where

f	c_V		c_A
$\nu_e,\, \nu_\mu,\, \nu_\tau$	0.5	$= \frac{1}{2}$	$\frac{1}{2}$
$e,\, \mu,\, \tau$	-0.0806	$= -\frac{1}{2} + 2\sin^2\theta_W$	$-\frac{1}{2}$
$u,\, c,\, t$	0.2204	$= \frac{1}{2} - \frac{4}{3}\sin^2\theta_W$	$\frac{1}{2}$
$d,\, s,\, b$	-0.3602	$= -\frac{1}{2} + \frac{2}{3}\sin^2\theta_W$	$-\frac{1}{2}$

This may seem complicated until one realizes that everything is controlled by $\theta_W = 28.75°$ which is called the Weinberg angle, that the relations

$$g_W = \frac{g_e}{\sin\theta_W} \qquad g_Z = \frac{g_e}{\sin\theta_W \cos\theta_W} \qquad m_W = m_Z \cos\theta_W$$

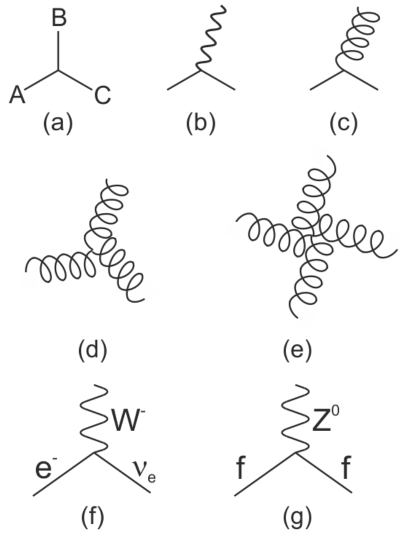
link the electromagnetic coupling constant with the ones of the weak interactions and that the masses of the gauge bosons of the weak force are connected as well. The vertex factors for $W^\pm$ and Z^0 are very similar but the coupling constants depend differently on the Weinberg angle. In the end all of this ugliness and the interconnectedness stems from the spontaneous breaking of the electroweak symmetry of the early universe where all the bosons were massless. There was only one coupling because there was one gauge group.

5.5 Calculations with Weak Interactions

Every particle that decays does it through the weak interactions. In calculations the external state labels look similar to Quantum Electrodynamics. There is no color, but despite the similarity there is a difference. When one labels external weak boson states with ϵ_μ and ϵ^*_μ they have three polarization states due to their mass while massless photons and gluons only have two. As shown above when discussing the Proca equation, internal propagators for massive particles look different than those for massless propagators because the change from massive to massless is discontinuous and one cannot simply set $m = 0$. The $W^\pm$ are antipartners and Z^0 is its own antiparticle.

The comparison of the different forces can be extended:

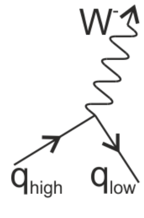
	External State Labels	Internal Propagators	Vertex Factors	
ABC	none	$\frac{i}{q^2 - m^2c^2}$	$-ig$	(a)
QED	$u,\, \bar{u},\, v,\, \bar{v}$	$\frac{i(\not{q} + mc)}{q^2 - m^2c^2}$	$ig_e\gamma^\mu$	(b)
	$\epsilon_\mu,\, \epsilon^*_\mu$	$-\frac{i\eta_{\mu\nu}}{q^2}$		
QCD	$uc,\, \bar{u}c^\dagger,\, vc,\, \bar{v}c^\dagger$	$\frac{i(\not{q} + mc)}{q^2 - m^2c^2}$	$-\frac{ig_s}{2}\lambda^\alpha\gamma^\mu$	(c)
	$\epsilon_\mu a^\alpha,\, \epsilon^*_\mu a^{\alpha *}$	$-\frac{i\eta_{\mu\nu}\delta^{\alpha\beta}}{q^2}$	$-g_s f^{\alpha\beta\gamma}[...]$	(d)
			$-ig_s^2[...]$	(e)
Weak	$u,\, \bar{u},\, v,\, \bar{v}$	$\frac{i(\not{q} + mc)}{q^2 - m^2c^2}$	$-\frac{ig_W}{2\sqrt{2}}\gamma^\mu(1 - \gamma^5)$	(f)
	$\epsilon_\mu,\, \epsilon^*_\mu$	$-\frac{i(\eta_{\mu\nu} - \frac{q_\mu q_\nu}{m^2c^2})}{q^2 - m^2c^2}$	$-\frac{ig_Z}{2}\gamma^\mu(c_V^f - c_A^f\gamma^5)$	(g)



where the ABC theory is a toy theory but no gauge theory, QED is Quantum Electrodynamics, QCD is Quantum Chromodynamics, and Weak is the theory of the weak interactions.

Finally, because the strong and weak operators do not commute, they have different eigenstates. Quarks are usually created in the eigenstates of the strong force but decay by the weak force and do it in eigenstates of the weak force

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix} \Rightarrow \begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix} \qquad \begin{pmatrix} q_\text{high} \\ q_\text{low} \end{pmatrix} \qquad -\frac{ig_W}{2\sqrt{2}}\gamma^\mu(1 - \gamma^5)V_\text{high-low}$$



where the CKM-matrix (5.1) plays its role.

In order to better understand the factor $\gamma^\mu(1-\gamma^5)$ appearing in the vertex factors, one looks at parity which is a reflection of an odd number of coordinates and is a discrete symmetry. Spinors transform under parity $P(x,y,z) \to (-x,-y,-z)$ as $\Psi \to \Psi' = \gamma^0\Psi$ which means that the simplest spinor sandwich

$$\bar{\Psi}\Psi \to \bar{\Psi}'\Psi' = \Psi'^\dagger\gamma^0\Psi' = (\gamma^0\Psi)^\dagger\gamma^0\gamma^0\Psi = \Psi^\dagger\gamma^{0\dagger} I\Psi = \Psi^\dagger\gamma^{0\dagger}\Psi = \Psi^\dagger\gamma^0\Psi = \bar{\Psi}\Psi$$

is a scalar and invariant. On the other hand the expression $\bar{\Psi}\gamma^5\Psi$ transforms as

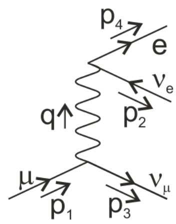
$$\bar{\Psi}\gamma^5\Psi \to \Psi^\dagger\gamma^{0\dagger}\gamma^0\gamma^5\gamma^0\Psi = \Psi^\dagger I\gamma^5\gamma^0\Psi = \Psi^\dagger\gamma^5\gamma^0\Psi = -\Psi^\dagger\gamma^0\gamma^5\Psi = -\bar{\Psi}\gamma^5\Psi$$

and is a pseudo-scalar. Similarly, one can show for the spinor sandwich $\bar{\Psi}\gamma^\mu\Psi \to -\bar{\Psi}\gamma^\mu\Psi$ which is a vector and for $\bar{\Psi}\gamma^\mu\gamma^5\Psi \to \bar{\Psi}\gamma^\mu\gamma^5\Psi$ which is a pseudo-vector or axial-vector.

This is interesting because if one constructs weak interaction Feynman diagrams one encounters spinors $\bar{u}\gamma^\mu(1-\gamma^5)u$ which is $\bar{u}\gamma^\mu u - \bar{u}\gamma^\mu\gamma^5 u$ and therefore a vector minus an axial-vector. (Note that the labels V and A in c_V^f and c_A^f reflect this, and that the corresponding parts in the vertex factor are called vector-part and axial-part of the coupling.) The significance of a vector minus an axial-vector is that it breaks parity, but the fact that weak interactions break parity is known since they only ever produce left-handed neutrinos. The factor $\gamma^\mu(1-\gamma^5)$ is therefore the source of the P-violation[2].

5.6 The Decay of a Muon

The muon is the second lightest lepton and decays as $\mu^- \to \nu_\mu + e^- + \bar{\nu}_e$ to an electron (and some particles due to conservation laws) because heavier particles decay to lighter particles. The amplitude becomes after some steps



$$\begin{aligned} M &= \bar{u}(3)\frac{-ig_W}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)u(1)\frac{-i(\eta_{\mu\nu}-\frac{q_\mu q_\nu}{m_W^2c^2})}{q^2-m_W^2c^2}\bar{u}(4)\frac{-ig_W}{2\sqrt{2}}\gamma^\nu(1-\gamma^5)u(2) \\ &\approx \frac{g_W^2}{8m_W^2c^2}\bar{u}(3)\gamma^\mu(1-\gamma^5)u(1)\bar{u}(4)\gamma_\nu(1-\gamma^5)u(2) \end{aligned}$$

where the approximation assumes $q \ll m_W c$ making the vertex factor approximately $\frac{i\eta_{\mu\nu}}{m_W^2c^2}$ which is typical for low-energy muons.

As usual one can average over the incoming spins and sum over the final spins using Casimir's trick and the γ matrix trace theorem to get

$$\langle|M|^2\rangle = 2\left(\frac{g_W}{m_W c}\right)^4 (p_1\cdot p_2)(p_3\cdot p_4)$$

in first order. To evaluate the muon lifetime as seen in its rest frame one uses p_1 with only the first non-zero component $m_\mu c$ and obtains

$$\langle|M|^2\rangle = \left(\frac{g_W}{m_W c}\right)^4 m_\mu^2 E_2(m_\mu c^2 - 2E_2)$$

after some work. To get the lifetime one applies Fermi's golden rule for decays (2.2), but this is a three-body decay and not a two-body decay such that one cannot use the simpler version (2.4). Working with $m_e \approx 0$ since $m_e c^2$ is a small percentage of the energy $(m_\mu - m_e)c^2$ basically released by a muon decay with $m_\mu \approx 105\,\mathrm{MeV}/c^2$ and $m_e \approx 0.5\,\mathrm{MeV}/c^2$ gives

$$d\Gamma = \frac{\langle|M|^2\rangle}{2\hbar m_\mu}\frac{d^3\vec{p}_2}{(2\pi)^3 2|\vec{p}_2|}\frac{d^3\vec{p}_3}{(2\pi)^3 2|\vec{p}_3|}\frac{d^3\vec{p}_4}{(2\pi)^3 2|\vec{p}_4|}(2\pi)^4\delta^4(p_1-p_2-p_3-p_4)$$

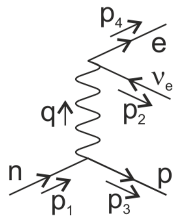
and in good approximation because this decay channel is by far the most probable

$$\tau = \frac{1}{\Gamma} = \left(\frac{m_W}{m_\mu g_W}\right)^4 \frac{12\hbar(8\pi)^3}{m_\mu c^2}$$

after some work. Using the observed muon lifetime $2.1970\cdot 10^{-6}\,\mathrm{s}$ and $m_W = 80\,420\,\mathrm{MeV}/c^2$ allows to calculate $g_W = 0.653$ and $\alpha_W = g_W^2/4\pi = \frac{1}{29.5}$. Compared to $\alpha_E = \frac{1}{137}$ for Quantum Electrodynamics shows that the weak force is not that weak, but that the $W^\pm$ are heavy.

[2]To understand the CP-violation, one has to dig into the theory related to the CKM-matrix.

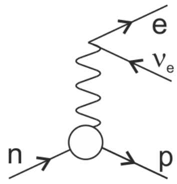
5.7 The Decay of a Neutron



The decay $n \to p + e + \nu_e$ and the corresponding Feynman diagram is a simplification because both neutron and proton are bound states of three quarks, but one can see how far one comes with this reduced complexity. The simplified Feynman diagram looks the same as the muon decay, and one can just take the amplitude

$$\langle |M|^2 \rangle = 2 \left(\frac{g_W}{m_W c} \right)^4 (p_1 \cdot p_2)(p_3 \cdot p_4)$$

calculated for the muon decay in first order. This time when employing the golden rule, one cannot ignore the contribution from m_e to the energy released $(m_n - m_p - m_e)c^2$ with $m_n = 939.6\,\mathrm{MeV}/c^2$, $m_p = 938.3\,\mathrm{MeV}/c^2$ and $m_e = 0.5\,\mathrm{MeV}/c^2$. This complicates the calculations considerably but gives after some work $\tau = \frac{1}{\Gamma} \approx 1318\,\mathrm{s}$ using the value of g_W found from the muon decay. The order of magnitude is not bad but the value is wrong since the value determined experimentally is $885.7\,\mathrm{s}$.

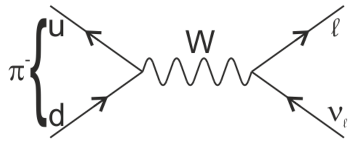


One can put in an unknown vertex. Because of energy-momentum conservation, electric charge conservation and so on, there is still some knowledge available. One can replace the factor $(1 - \gamma^5)$ by $c_V - c_A\gamma^5$ with the lacking knowledge in the two numbers c_V and c_A. Experimentally one finds $c_V \approx 1$ and $c_A \approx 1.27$. The vector value c_V means weak vector charge conservation. Thus, the complications in the unknown vertex had little influence on it, but it changes the axial part. The fact that the vector part does not change is called the conserved vector current hypothesis or shorter CVC hypothesis.

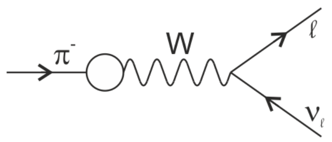
To be fair, it is hard to say just how W couples to the n or p since they are a "hot QCD mess" of quarks, virtual gluons and virtual quark-antiquark pairs. This is not an issue when a photon couples to the hot QCD mess since even in a mess of virtual pairs, the net electric charge is conserved. There is no theoretical starting point for "weak charge" conservation. So a similar calculation with Quantum Electrodynamics where two photons exchange a photon gets a much better answer.

Just using the diagram where two protons exchange a photon gives pretty much the right result for the scattering amplitude of two protons. However, the same problems with the hot QCD mess should occur here because each proton consists of three quarks and it is not clear which of these quarks exchange the photon. All the same complications as for the neutron decay could exist. There is a symmetry because whatever is going on, the total electric charge is conserved. In the case of the neutron decay there is no similar conservation which protects the amplitude of the photon exchange between two protons.

5.8 The Decay of a Pion



The decay of a pion into a lepton and the corresponding antineutrino as in the process $\pi^- \to l + \nu_l$ is different from the other two decays because the two quarks are annihilating to form a W boson. It is an exact diagram and not a simplification as the diagram in the neutron decay. This diagram is very hard to evaluate because the quark and antiquark are bound to each other and it is not just two free particles. The calculation of the π^- lifetime is challenging since it depends on the ground state wave function of the system consisting of the down quark and the up antiquark. What can be independently calculated is the ratio of $\pi^- \to e^- + \bar{\nu}_e$ to $\pi^- \to \mu^- + \bar{\nu}_\mu$.



Similarly to the neutron decay above one can add a bubble of ignorance and calculate the amplitude as

$$M = \frac{g_W^2}{8(m_W c)^2} \bar{u}(3)\gamma_\mu(1 - \gamma^5)v(2)F^\mu$$

where F^μ encodes the ignorance in the bubble. This F^μ must depend on some 4-vector available before the creation of the lepton-neutrino pair to make M a scalar, but the only one is the momentum P^μ of the incoming π^- such that $F^\mu = f_\pi P^\mu$. The scalar f_π must depend on the only scalar available $P^2 = m_\pi^2 c^2$. This means that the result for F^μ is independent of which lepton and corresponding neutrino is used.

Thus one can calculate the ratio of the decay into the electron and the decay into the muon, because this factor of ignorance cancels out. One gets

$$\langle |M|^2 \rangle = \frac{1}{8}\left[f_\pi \left(\frac{g_W}{m_W c} \right)^2 \right]^2 \left[2(p \cdot p_2)(p \cdot p_3) - p^2 (p_2 \cdot p_3) \right] \qquad \Gamma = \frac{f_\pi^2}{\pi \hbar m_\pi^3} \left(\frac{g_W}{4 m_W} \right)^4 m_l (m_\pi^2 - m_l^2)^2$$

using the momentum p of the incoming pion because one can use the two-body decay formula (2.4) this time and $|\vec{p}_2| = \frac{c}{2m_\pi}(m_\pi^2 - ml^2)$.

One cannot evaluate Γ without f_π but one can calculate the ratio of the decay into the electron channel versus the decay into the muon channel

$$\frac{\Gamma(\pi^- \to e^- + \bar{\nu}_e)}{\Gamma(\pi^- \to \mu^- + \bar{\nu}_\mu)} = \frac{m_e (m_\pi^2 - m_e^2)^2}{m_\mu (m_\pi^2 - m_\mu^2)^2} = 1.283 \cdot 10^{-4}$$

which is surprising because one would expect that the kinematic likelihood is driven by the mass difference and that the decay into an electron is more likely than the decay into a muon. The mass difference between a pion and an electron is larger than the mass difference between a pion and a muon. One of the driving factors behind decay rates or any other process is the amount of phase space available. If the mass difference between the pion and the lepton is large then there is a lot of energy to distribute in various ways. Therefore the decay into an electron should be more probable than the decay into a muon.

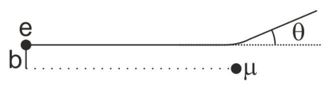
The reason for this counterintuitive fact is that the π^- is a spin-0 particle so when it decays the lepton l and the corresponding neutrino ν_l must come out with equal helicity and opposite spins. As a two-body decay they have to come out with back-to-back momentum if the pion is at rest. In the universe there are only right-handed antineutrinos, and the spin of the outcoming antineutrino must therefore point into the direction of the momentum. The weak interactions produce left-handed massless matter and right-handed massless antimatter. Thus, because m_e is nearly zero it is almost always produced as a left-handed particle whereas m_μ is very massive and so happy to exist as a right-handed particle. The argument of the mass difference therefore does not work for the probability of the decay into an electron versus the decay into a muon, but it works for the decay into a muon versus the decay into a tauon.

6 Renormalization

6.1 Motivation for Renormalization

Renormalization is what is used to connect Particle Physics to other areas of physics. Its importance goes way beyond just Particle Physics because it also ties other areas of physic together.

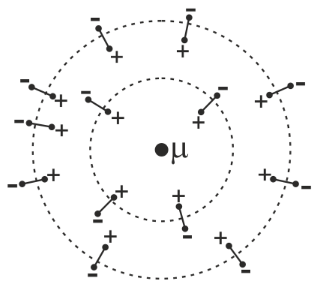
In classical physics a muon is in the following assumed to be at rest. If one wants to measure its charge q_μ, one might scatter a known charge with known momentum P^μ and use the observed trajectory to infer q_μ. If an electron is used, one could even work in the limit that the muon remains at rest the whole time. If the impact factor b is known, one can use



$$q_\mu = \frac{2bE}{q_e \cot\left(\frac{\theta}{2}\right)}$$

which is a formula from classical scattering theory.

Now the muon is dropped into a dielectric, and the dielectric is responding to this negatively charged intruder. Each of the charged dipols has one positive and one negative end, and the positive end is attracted by the muon such that the muon will gently reorient the dipols. The effective charge in the two Gaussian surfaces in the figure are different. For the inner Gaussian surface there is $q_\mu + 4e$ and for the outer there is $q + 9e$. The four dipols with both ends completely inside the outer Gaussian surface do not contribute. Thus, the effective charge varies with distance.

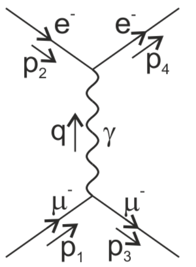


If one repeats the scattering experiment with the muon in the dielectric and sends in a probe electron, it will only respond to the effective charge which varies with distance. The result of this experiment is very sensitive to how close the electron comes to the muon. In particular, the higher the kinetic energy of the electron the closer it comes to the muon. Thus, an electron with lower energy will see a greater screening, and an electron with higher energy will see a smaller screening. If one calls the distance of closest approach $\tilde{r}$, the the largest value seen would be $q_\mu(\tilde{r})$ but this is also a function of the energy such that $q_\mu(\tilde{r}) \sim q_\mu(E)$. (It is assumed that the impact parameter is fix.)

There is this dielectric and the charge of the muon is supposed to be measured, but what one really measures is the effective charge which scales with energy. However, there is a way to measure the real charge without screening. One just has to come closer to the muon than any screening pairs with sufficiently high energy (or one can remove the muon from the dielectric).

The quantum picture of this situation (where definite results are replaced with probabilities, and where the particles are fluctuations of the corresponding field) is as if the electron feels the electromagnetic field of the muon, but the mechanism is the exchange of a photon. That is not the only Feynman diagram, because the photon can split into a virtual electron-positron pair which annihilates and becomes a photon again. The virtual electron-positron pair between the electron and the muon could act as a dielectric such that they might screen the charge of the muon. This classical story of screening and effective charge when it is rolled into the quantum version is essentially the story of renormalization in Quantum Field Theory.

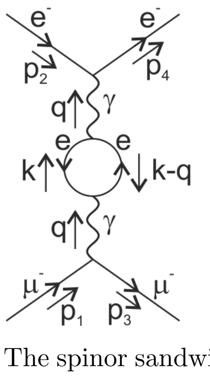
6.2 Screening Effects in Quantum Electrodynamics



Exploring the screening of real electric charge by virtual particle pairs in a more quantitative way will lead to a better understanding for divergences encountered in Feynman amplitudes. The scattering process $e^- + \mu^- \to e^- + \mu^-$ has been studied above in the context of Quantum Electrodynamics and will here be studied in the context of renormalization.

The simplest non-trivial Feynman diagram is the second order diagram which is also called "tree-level" diagram because trees usually have no holes. Its contribution to the amplitude is according to (3.3)

$$M_{\text{tree}} = -g_e^2 \big[\bar{u}(3)\gamma^\mu u(1)\big] \frac{\eta_{\mu\nu}}{q^2} \big[\bar{u}(4)\gamma^\nu u(2)\big]$$



where $q = p_2 - p_4$ is the momentum transfer.

The effects of virtual particle pairs start at fourth order with the largest contribution from a virtual electron-positron pair as in the figure where momentum conservation simplified labeling the internal momenta. The contribution to the amplitude is

$$M_{\text{loop}} = -\frac{ig_e^4}{q^4} [\bar{u}(3)\gamma^\mu u(1)] \left[\int \frac{d^4k}{(2\pi)^4} \frac{\mathrm{Tr}\big[\gamma_\mu(\not{k} + m_e c)\,\gamma_\nu(\not{k} - \not{q} + m_e c)\big]}{[k^2 - m_e^2c^2][(k-q)^2 - m_e^2c^2]} \right] [\bar{u}(4)\gamma^\nu u(2)]$$

according to (3.6) also with $q = p_2 - p_4$.

The spinor sandwiches in M_{tree} and M_{loop} are the same. The integral in M_{loop} however is divergent. (The integral over d^4k can be seen as an integral over $k^3\,dk$ and with counting the factors k in the numerator and the denominator gives an integral over $k\,dk$ from zero to plus infinity because k can be arbitrarily large. More careful analysis shows that it is only a logarithmic divergence, but it is still a divergence.) One encounters this kind of infinities all over the place in Quantum Field Theory when going beyond leading order. In the sixties these infinities caused great headaches and some scientists thought that Quantum Field Theory should be abandoned. Some progress had been made but the solutions were not really convincing.

Calling the integral in (3.6) between the two spinor sandwiches $-i\hbar\eta_{\mu\nu}I(q^2)$ gives

$$I(q^2) = -\frac{1}{12\pi^2} \left\{ \int_{m_e^2}^{\infty} \frac{dz}{z} - 6\int_0^1 z(1-z)\ln\left[1 - \frac{q^2}{m_e^2c^2} z(1-z)\right] dz \right\}$$

after long and heavy calculations. The important points are that the first integral which does not depend on q diverges as $\ln z$, and that the second integral which depends on q is finite. The q is the momentum exchange, and this momentum exchange depends on how hard the electron is sent to the muon. The second integral results in a function

$$f\left(\frac{-q^2}{m_e^2 c^2}\right)$$

with $f(-q^2 = 0) = 0$ and $f(-q^2)$ increases with decreasing q^2. (Note that $q^2 < 0$ for this process.)

In the program of renormalization which is the discovery of infinities and doing something with them, the first step is finding an infinity as this one. The second step is regularizing the divergence and make it a finite contribution. This is kind of cheating because one does not integrate to infinity but just to a finite upper limit called a cut-off. (There are other ways to make infinite contributions finite.) Here the cut-off is m_c^2 and all integrals

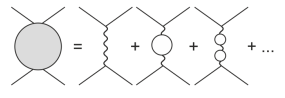
$$\int_{m_e^2}^{m_c^2} \frac{dz}{z} = \ln\left(\frac{m_c^2}{m_e^2}\right) \qquad\qquad I(q^2) = -\frac{1}{12\pi^2}\left\{\ln\left(\frac{m_c^2}{m_e^2}\right) - f\left(\frac{-q^2}{m_e^2 c^2}\right)\right\}$$

become finite. The amplitude is

$$M_{\text{tree}} + M_{\text{loop}} = -g_e^2[\bar{u}(3)\gamma^\mu u(1)]\frac{\eta_{\mu\nu}}{q^2}\left[1 - \frac{g_e^2}{12\pi^2}\left\{\ln\left(\frac{m_c^2}{m_e^2}\right) - f\left(\frac{-q^2}{m_e^2 c^2}\right)\right\}\right][\bar{u}(4)\gamma^\nu u(2)]$$

where the divergence is still present when $m_c \to \infty$.

6.3 Interpretation of the Screening Effects



One does not measure M_{tree} or M_{loop} but the total M which contains the tree-level diagram and all the higher order diagrams with more and more loops among other things. Thus one does not measure the true charge but the effective charge with the screening effects of the higher order diagrams coming from the loop contributions when one interprets the measurement classically as if there is only the tree-level diagram. To formalize this one simply rewrites the results in terms of the effective charge and coupling. This renormalized coupling constant is

$$g_R(q^2) = g_e\sqrt{1 - \frac{g_e^2}{12\pi^2}\left[\ln\left(\frac{m_c^2}{m_e^2}\right) - f\left(\frac{-q^2}{m_e^2 c^2}\right)\right]}$$

and is the physically measured value of $e\sqrt{4\pi/\hbar c}$ in experiments at momentum transfer scale $q = p_2 - p_4$.

The amplitude can now simply be written as

$$M_{\text{tree}} + M_{\text{loop}} = -g_R(q^2)[\bar{u}(3)\gamma^\mu u(1)]\frac{\eta_{\mu\nu}}{q^2}[\bar{u}(4)\gamma^\nu u(2)]$$

which looks like coming from one simple tree-level diagram alone but contains all the contributions of the higher-order diagrams. However, this is not very convenient because g_R depends on q. The renormalized coupling is therefore often written as

$$g_R(q^2) = g_R(0)\sqrt{1 + \frac{g_R^2(0)}{12\pi^2} f\left(\frac{-q^2}{m_e^2 c^2}\right)} \qquad \text{with} \qquad g_R(0) = g_e\sqrt{1 - \frac{g_e^2}{12\pi^2}\ln\left(\frac{m_c^2}{m_e^2}\right)}$$

in terms of the zero-momentum (large distance) value $g_R(0)$. This is an example of what is called an effective theory. One considers a simpler looking theory where the quantum loop corrections are bundled into renormalized quantities and the result is interpreted as the classical tree-level diagram.

But the question remains, what are the three numbers $g_e = e\sqrt{4\pi/\hbar c}$, $g_R(0)$ and $g_R(q^2)$ and which of them does one measure as the electromagnetic coupling constant. If one looks up the coupling constant is a physics book one finds $g_R(0)$. This is the electromagnetic coupling when doing a macroscopic experiment because it is the value for the zero-momentum transfer. However, $g_R(0)$ is equal to g_e times an infinite factor for $m_c \to \infty$. The value g_e is called the bare or the fundamental electric coupling which would be the coupling if one could come so close to the muon that no screening effects interfere. In the discussion of the classical case there were two options to measure the electric charge of the muon. One can take the muon out of the dielectric or one can shoot the electron with so much energy that it comes closer to the muon than any of the dielectric dipols. One cannot do either of those two possibilities in Quantum Field Theory because it is not possible to take the particle out of the bath of those virtual loops, and if one wants to probe it to see the bare coupling one has to probe it with infinite energy.

Quantum Field Theory has been built based on the coupling g_e and therefore based on a value one does not have access to. Going to $g_R(0)$ means that the theory has been rewritten based on a value one can measure and compare to the results of experiments. Thus the fact that $g_R(0)$ is defined as g_e times an infinite factor is irrelevant because one does not know anything about g_e which could itself be an infinity. What one knows for sure is that $g_R(0)$ is finite.

The process of renormalization has two very important ingredients. The first point is that the typically arising infinities are associated with parameters one can never truly know. If one writes the theory in terms of quantities one can measure these infinities disappear. Such a theory is renormalizable. The second observation which has nothing to do with infinities is that coupling constants run with energy scale. They are therefore no constants.

Whether the little bubbles in the Feynman diagrams corresponding to virtual electron-proton pairs and so on are screening the electric charge or are they antiscreening is another question. In other words, one can ask whether $g_R(q^2)$ is larger or smaller than $g_R(0)$. It is increasing because f is increasing. This means that the closer one gets to the muon the larger is the charge of the muon, and therefore the bubbles are screening.

6.4 Renormalization in General

After using all delta function the amplitude is $M \sim \delta^4(p_{\text{tot in}} - p_{\text{tot out}}) \int f(p_i, q_k) \prod_k \frac{d^4 q_k}{(2\pi)^4}$ where the q_k are the leftover internal momenta q after using the delta functions. In many situation the remaining integral diverges – usually at large q_k. One can render the integral finite by regularizing it. One possibility is to introduce a factor $(-\tilde{m}^2c^2)/(q_k^2 - \tilde{m}^2c^2)$ called Pauli-Villars massive regulator which increases the power of q_k in the denominator. The quantity $\tilde{m}$ is a free parameter and is not related to m_k. It is an interesting way to write 1 because with $\tilde{m} \to \infty$ it becomes 1. If the integral breaks into two parts where one depends on $\tilde{m}$ and becomes infinite for $\tilde{m} \to \infty$ and the other part is independent of $\tilde{m}$ and remains finite, and if the part depending on $\tilde{m}$ looks like $g_p = g_b + \delta g(\tilde{m})$ where g_p is the measurable quantity, g_b is the bare quantity and only δg depends on $\tilde{m}$, then one can ignore g_b and the infinity because g_b may be diverging too such that it cancels the divergence of δg. Only the physically measurable value g_p is of interest. The problem arises when one tries to formulate the theory in terms of g_b instead of g_p. If rewriting with $g_R(0)$ works, the theory is called renormalizable. Any theory can be regularized to make it finite. The key to renormalizability is removing the regulator in a consistent way[3]. Not only the coupling but also the mass can diverge with energy scale such that the bare mass is impossible to know.

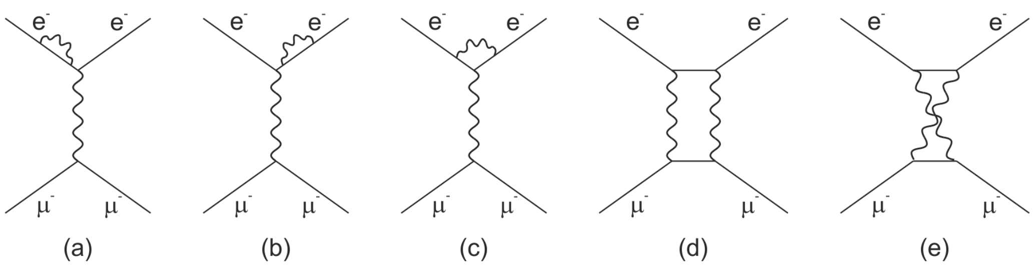


Figure 6.1: Fourth-order diagrams for electron-muon scattering

[3]Gerardus 't Hooft demonstrated the full renormalizability of the Standard Model, but perturbative quantum gravity (gravity as a gauge theory) unfortunately turned out to be non-renormalizable.

In the above considerations of electron-muon scattering in Quantum Electrodynamics only fourth-order diagrams have been taken into account where the photon exchanged between electron and muon split at least once into a virtual electron-positron pair. Diagrams with bubbles of electron-positron pairs did lead to infinities, but they could be handled by redefining g_e which is basically the electric charge of the electron. Some but not all of the other fourth-order diagrams are shown in figure 6.1. The diagrams (a), (b) and (c) lead to infinities, but (d) and (e) are finite and do not contribute to renormalization.

The electron self-energy corrections in (a) and (b) renormalize the electron mass m_e. The little photon loop does not care about the rest of the diagram, and it even does not care about the muon involved but only about the electron flying through space which emits and absorbs this virtual photon. It contributes non-trivially to the self-energy of the electron and therefore to the rest mass of the electron. The emerging infinities can be cured by renormalizing the electron mass m_e similarly to the electron charge g_e above. The vertex correction in (c) renormalizes the magnetic moment of the electron. (Similar diagrams would lead to self-energy corrections and vertex corrections for the muon.)

The three diagrams (a), (b), (c) separately contribute to the electron charge renormalization along with the diagrams containing loops of virtual electron-positron pairs replacing the simple electron exchange between electron and muon. However, the combined effect of these three cancels in Quantum Electrodynamics, and only the loops with virtual electron-positron pairs play a non-trivial role. This is good because the corrections to the charge from these three diagrams would be mass dependent leading to different effective charges for the three charged leptons. Since the same effective charges are observed, the contributions must cancel. By the way, the different parameters like mass m_e and charge g_e are not renormalized separately, but parameters need to be found such that all the infinitely many diagrams can be replaced by one single second-order diagram.

6.5 Renormalization in Quantum Chromodynamics

In Quantum Electrodynamics the vacuum polarization led to effective charge screening which made the electric charge run to larger values at smaller distances or at larger $|q^2|$ momentum transfer. Taking only the tree-level and the first loop diagram for electron-electron scattering into account gives

$$\alpha_e(|q^2|) = \alpha_e(0)\left[1 + \frac{\alpha_e(0)}{3\pi}\ \ln\left(\frac{|q^2|}{m_e^2c^2}\right)\right]$$

for $|q^2| \gg m_e^2c^2$ where $g_e = \sqrt{4\pi}\alpha_e$. Taking also the higher-order loops into account gives

$$\alpha_e(|q^2|) = \frac{\alpha_e(0)}{1 - \frac{\alpha_e(0)}{3\pi}\ \ln\left(\frac{|q^2|}{m_e^2c^2}\right)}$$

also for $|q^2| \gg m_e^2c^2$ because this is a geometric series. Note however that this does not count all possibilities. A loop where the virtual electron-positron pair exchanges a photon, for example, is not included. Only the leading terms for loops at each order are counted. The value $\alpha_e(|q^2|)$ increases with increasing $|q^2|$ and even blows up when the denominator gets zero but this happens at 10^{280} MeV and is therefore not too troubling. The fact that $\alpha_E(|q^2|)$ increases with increasing $|q^2|$ means screening.

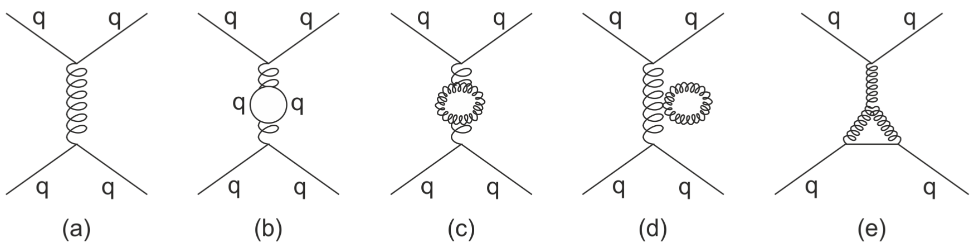


Figure 6.2: Higher-order diagrams for quark-quark scattering

In Quantum Chromodynamics quark-quark scattering by the exchange of a gluon allows more diagrams because of the gluon-gluon interactions. In figure 6.2 some possibilities are shown. The diagrams in

(a) and (b) correspond to the tree-level diagram and the loop diagram with a virtual quark-quark pair instead of the electron-positron pair of Quantum Electrodynamics. They lead to screening. There are also the diagrams (c) with a three-gluon vertex, (d) with a four-gluon vertex and more complex diagrams such as (e). They lead to antiscreening. Which effect wins depends on the number of different quark flavors f and the number of colors n which impacts both the number of quarks and gluons. A larger number of quarks leads to more screening and a larger number of gluons to more antiscreening.

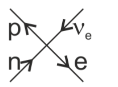
Evaluating loop diagrams in Quantum Chromodynamics requires Faddeev–Popov ghosts and is beyond the prerequisite knowledge here, but the result is

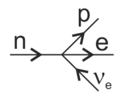
$$\alpha_S(|q^2|) = \frac{\alpha_S(m^2)}{1 + \dfrac{\alpha_S(m^2)}{12\pi}\,(11n - 2f)\,\ln\left(\dfrac{|q^2|}{m_e^2}\right)}$$

analogous to the Quantum Electrodynamics case. Similar to $\alpha_E(0)$ there $|q^2| \gg m^2$ is some scale at which one sets a reference value of α_S but one cannot use $\alpha_S(0)$ because it diverges. It turns out that if $11n - 2f > 0$ antiscreening wins, and if $11n - 2f < 0$ screening wins. In the Standard Model antiscreening wins and $\alpha_S(|q^2|)$ decreases with increasing $|q^2|$ because $n = 3$ and $f = 6$ resulting in $33 - 12 = 21 > 0$.

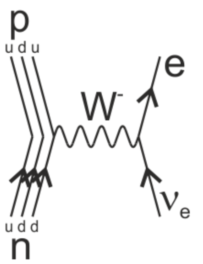
If one throws quarks together at very high energy their coupling is getting smaller and smaller. This means that at very high energies quarks are not interacting. This leads to asymptotic freedom which allows to effectively use perturbation theory in Quantum Chromodynamics at short distances like inside mesons and baryons. Insides mesons and baryons the quarks are barely interacting with each other. Of course one could try to extrapolate from this that at low $|q^2|$ or large distances the $\alpha_S(|q^2|)$ goes to infinity and hence explain confinement, but in this region perturbation theory breaks down because the coupling gets bigger than one[4] and this does not prove the confinement hypothesis.

6.6 Effective Field Theories





Before it was known that the weak force is based on a SU(2) invariance and the strong force on a SU(3) invariance, Fermi proposed a field theory model to describe the beta decay. His theory included four fermion fields p, n, e, ν_e and an interaction vertex called the Fermi four-point interaction. This could obviously describe a beta deday $n \to p + e + \bar{\nu}_e$ drawn with one arrow coming in and three going out, two in the direction of time and one opposite. Using this field theory one can calculate the amplitude for this process. Eventually this was proven to be wrong by experimental observation. Interestingly also the theory showed that this is not correct because it was not renormalizable.



In this case one can understand why this theory was broken. In truth the underlying process is the exchange of a W^- boson, and both the proton and the neutron consist of three quarks. In a sense the Fermi four-point interaction theory is collapsing the W^- line to zero. To see the conditions for when this is reasonable one can just consider the amplitude M which is without some factors

$$\left[\bar{u}(n)\gamma^\mu(1-\gamma^5)u(p)\right]\frac{-i(\eta_{\mu\nu} - \frac{q_\mu q_\nu}{m_W^2 c^2})}{q^2 - m_W^2 c^2}\left[\bar{v}(\nu_e)\gamma^\nu(1-\gamma^5)u(e)\right]$$

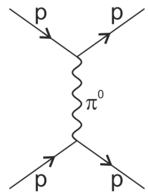
for the process as seen today. This is a vertex factor times a propagator for the weak interaction times a vertex factor. Fermi's theory of the beta decay breaks when q gets too big. For $q \ll m_W c$ the part of the amplitude M above becomes

$$\left[\bar{u}(n)\gamma^\mu(1-\gamma^5)u(p)\right]\frac{i\eta_{\mu\nu}}{m_W^2 c^2}\left[\bar{v}(\nu_e)\gamma^\nu(1-\gamma^5)u(e)\right]$$

showing that for $q \ll m_W c$ the four-point interaction is useful, but when $q \gtrapprox m_W c$ the full underlying theory is needed. This is a vertex factor times a vertex factor without the propagator.

[4]As a side note, theoretists like to play with "pure" Quantum Chromodynamics with only gluons since it is actually a finite theory which does not need renormalization.

Another example of an effective field theory which had to be replaced by a refined theory is the very successful model of the strong interactions before Quantum Chromodynamics in terms of matter fields for the proton and neutron and gauge fields corresponding to the pions with fundamental vertices. It was an SU(2) gauge theory acting on nucleon doublets with a neutron and a proton and with a gauge field π^0, $\pi^{\pm}$ for each generator of SU(2). Calculating processes such as $p+p \to p+p$ with this theory was much easier than working with Quantum Chromodynamics.



The lesson here is that if one writes down a non-normalizable theory, it is not that it is physically meaningless but it means that this is an effective theory and has to be replaced by a more fundamental description when going to high energies. The more fundamental description is typically called the ultraviolet completion of the theory where ultraviolet means high energy. The expectation is, of course, that the ultraviolet completion itself is either renormalizable or – even better – final such that one does not have to mess with renormalization because there are no infinities in the theory.

6.7 Wilson's Approach to Renormalization

In the seventies Kenneth Wilson formalized the concepts around renormalizable and non-renormalizable theories. The basic idea is that a non-renormalizable theory with unremovable infinities is really an effective theory that should only be used up to some energy scale. Beyond that limit one should instead work with the more fundamental ultraviolet completion of the theory.

Turning this around, Wilson introduced the notion of moving from fundamental descriptions to effective theories by "integrating out" the higher energy degrees of freedom, and working only in terms of the lower energy degrees of freedom. This usually leads to a simpler type theory. Note that this is exactly what has been done to get the Fermi model by ignoring the highest energy and therefore most massive part of the fundamental description. The mass of the W boson is by far the biggest thing in the beta decay because the quarks and the leptons are trivial in mass compared to the W. Thus, ignoring W is removing the highest energy degree of freedom from the theory.

In the example of the Fermi four-point interaction and the more fundamental description in terms of quarks and the W boson, both theories were field theories, but this program does not have to connect field theories to field theories. For example, one could approximate a discrete atomic system at large distance by an approximate continuum description. This effective field theory description would be non-renormalizable and would break down at energy scales associated with the atomic spacing. One does not expect this theory to be good at arbitrarily high energies or small distances.

6.8 Quantum Gravity

This leads back to the problem child perturbative quantum gravity where General Relativity is taken to be worked with in analogy to the way one works in the Standard Model. It is non-renormalizable, and the current interpretation is that it is still a useful effective field theory that should be replaced at some appropriate energy scale by it ultraviolet completion. If one takes the fundamental constants $\hbar$, c and the gravitational constant G, then one can form

$$m_{\text{Planck}} = \sqrt{\frac{\hbar c}{G}} \approx 10^{-8}\,\text{kg} \qquad E_{\text{Planck}} = m_{\text{Planck}} c^2 \approx 10^{19}\,\text{GeV} \qquad l_{\text{Planck}} = \sqrt{\frac{\hbar G}{c^3}} \approx 10^{-35}\,\text{m} \tag{6.1}$$

which are called Planck units. (The Planck mass is huge compared to the mass of the heavy Higgs or the equally heavy W boson in the order of 10^{-25} kg). One can rely on perturbative quantum gravity up to very high energies and short distances, but if one wants an accurate description of gravity very near the big bang or the singularity of a black hole an ultraviolet completion is needed.

The two possible candidates for this ultraviolet completion of perturbative quantum gravity are:

1. Loop Quantum Gravity: It is a straightforward quantization of General Relativity and addresses the problem of quantum gravity more than it solves it.
2. String Theory: It handles quantum gravity and much more, and it comes equipped with supersymmetry crucial for gauge force unification.

Appendix A: Some General Topics

Baryon Asymmetry of the Universe

The question is why there is so much matter and so little antimatter. When discussing what "matter" fills the universum, for practical purposes it is sufficient to count the number of barions (three-quark bound states) which occur largely in the form of protons and neutrons while other baryons are unstable. Of course the total matter in the universe is largely electrically neutral so it is known that for every proton there must exist a negatively charged particle, but these are usually electrons whose mass is tiny compared with the proton or neutron mass and makes them relatively inconsequential. By the way, one only needs to be concerned with electrons, protons and neutrons because these are long lived as also are neutrinos which however are even less consequential than electrons. This also means that one does not need worry about any mesons which are unstable. There are about 10^{80} baryons and about 0 antibaryons in the universe, and this is a surprising asymmetry.

One reason why the baryonic asymmetry is so interesting is that it is in reality a surprisingly small difference between baryons and antibaryons because one should not think about the universe now but about the universe very early on. If one traces back through the confidently understood part of the history of the universe, one finds that at about a billionth of a second after the big bang that the difference was actually only one part per billion. That is, at that time, the universe was tiny but had the same stuff in it as today such that it was very hot and the energy density was extremely high. With this energy density all kinds of pairs of particles and antiparticles were created, and what one finds at this time-frame was that for every billion antibaryons in the universe there was one billion plus one baryon. Since then the universe cooled down and the billions of baryons and antibaryons have annihilated, and what we see today is that one part per billion left over.

This fact makes the baryonic asymmetry much harder. It is much easier in physics to explain zero than it is to explain a small number. Symmetry is an easy argument for a difference being zero. The challenge is to think about Particle Physics in the context of Cosmology including General Relativity and come up with an explanation for the baryonic asymmetry. Physicists believe that the universe was created symmetrically but became asymmetric through some later process. The challenge is therefore to find a scenario where this asymmetry was generated later starting from an equal number of baryons and antibaryons. Something in the early universe must have caused these two numbers to become different.

In 1967 Sakharov enumerated three conditions that would have to be met for this to even be possible:

1. Baryon number violation
2. C and CP violation
3. Departure from thermal equilibrium

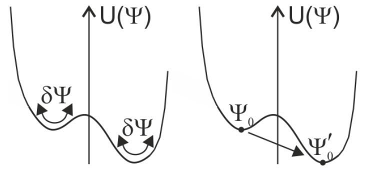
Their relation to the baryonic asymmetry is discussed in the following.

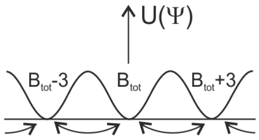
Baryon Number Violation

The first condition is baryon number violation. The baryon number B_{tot} is defined as the total additive number when each baryon is counted $+1$ and each antibaryon is counted -1. A symmetric universe has $B_{\text{tot}} = 0$ and the current universe has $B_{\text{tot}} \approx 10^{80}$. Conservation of a quantity does not mean that it is constant but means that one can explain if this quantity changes in a system by a flux through the boundary of this system. If the system to be taken is the entire universe for which there is no way for things to enter or exit, then baryon number conservation means that the total number of baryons is constant. However, if the universe was born with $B_{\text{tot}} = 0$ and got $B_{\text{tot}} \approx 10^{80}$ in the time since then baryon number conservation must be violated.

If one takes the Standard Model Lagrangian and all of the Feynman diagram vertices that it predicts, one discovers that it is impossible to draw a diagram that connects a set of incoming states to a set of outgoing states where the baryon number for the incoming states is not the same as the baryon number of the outgoing states. Hence, the Standard Model conserves baryon number at least pertubatively. The Feynman rules are a perturbative treatment of small fluctuations. They apply to perturbations $\delta\Psi$ and not to the original field Ψ which is approximated as $\Psi \rightarrow \Psi_0 + \delta\Psi$.

Given a potential $U(\Psi)$ with two local minima and a local maximum in between small fluctuations of $\delta\Psi$ can be described by Feynman diagrams at the two minima, but tunneling from Ψ_0 to Ψ_0' is not possible because tunneling is not perturbative. A small fluctuation is stuck in a valley, and moving over the maximum is not a small fluctuation. With the Higgs mechanism there are non-perturbative phenomena in the Standard Model. The Higgs mechanism itself is not able to create baryons, but there may be non-perturbative mechanisms which do.





It turns out that in the Standard Model Lagrangian a non-perturbative effect that violates baryon number does arise. With electromagnetism and the strong interaction it does not happen, but with the weak force or the electroweak force in the early universe there is a series of minima with degenerate energy. The important fact about these different possible minima is that they differ in the total number of baryons. It turns out for the specific Standard Model Lagrangian process that the difference is by three baryons. It actually conserves the baryon number minus the lepton number because if it creates three baryons it also creates three antileptons, and if it creates three antibaryons it also creates three leptons. The different minima or vacua represent different configurations of the field Ψ.

The arrows from one vacuum to the next point in both directions. This means that the baryon number can be increased and decreased. The transition from one vacuum to the next is either tunneling where the process is called instantan or jumping over the bump where the process is called sphaleron. In the early universe with the high energy density there was a lot of jiggling such that jumping over the bumps could happen. These sphaleron processes are part of the electroweak sector of the Standard Model Lagrangian. It is not completely clear whether the process shown here is the reason for the 10^{80} baryons, but there is at least one known process which violates the conservation of the baryon number.

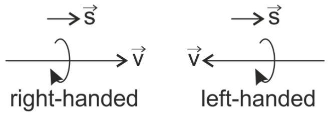
C and CP Violation

One big problem is that the arrows from one vacuum to the next, and the process to increase the baryon number has the same amplitude as the process to decrease the baryon number. To come to 10^{80} baryons a mechanism is needed to force the process to only go in the direction which increases the baryon number. One way to relate the $B \to B + 3$ process to the $B \to B - 3$ process is through charge conjugation C. This transformation simply interchanges all particles with their antiparticle versions and vice versa for a process. However, despite its name it does more because it also interchanges the electrically uncharged neutrinos by their antiparticle versions. Charge conjugation is certainly a symmetry which is satisfied by many processes in Particle Physics. If one calculates the scattering amplitude for a process with two electrons coming in and two muons going out and the scattering amplitude for the process where all particles are replaced by their antiparticles, then the two scattering amplitudes are the same. The Feynman diagrams for both processes look the same, and the masses are all the same.

There are other processes where charge conjugation is not a symmetry. The pion Π^+ decays into an antimuon and a muon-neutrino. If the Π^+ starts at rest, the two outgoing particles have to emerge back to back. The Π^+ is a spin-0 particle, and the spins of the outgoing particles which are both spin-$\frac{1}{2}$ have to be anti-aligned, and they both have to be left-handed. This process happens a lot, and one can calculate the amplitude M which is not zero. Calculating the charge conjugate shows a Π^- which decays into an antimuon and an antimuon-neutrino. The total momentum of the Π^- is still zero. The two outgoing particles emerge back to back and they would still have the same spin. Calculating the amplitude for this process gives zero, and one does never see this happening. This is an example which violates C maximally.

If one imagines that this effect also carries over to the electroweak spaleron processes which violate B, then one can envision that the process of adding three baryons has a non-zero amplitude while the process of adding three antibaryons has a zero amplitude because the process $B \to B - 3$ is the charge conjugate of the process $B \to B + 3$. This would not violate charge conservation but charge conjugation. It is not necessary that one amplitude is zero, but the probability for the two processes must be different. There are charge conjugated processes in the Standard Model with different frequencies.

However, while one does not observe the charge conjugate version of the Π^- decay, one does observe the version where charge and parity are conjugated. Parity is, among other things, responsible for inverting handedness. CP is a combination of charge conjugation and parity transformation. (Note that parity is the transformation which has been removed when going from O(3) to SO(3), for example.) The momentum changes direction but the spin does not through a parity transformation. Momentum is a true vector while any kind of angular momentum (including spin) is a pseudo-vector.



The CP-conjugated version of the Π^- decay into a left-handed antimuon and a left-handed antimuon-neutrino has the same amplitude as the decay of a Π^+ into a right-handed muon and a right-handed muon-neutrino. This is a problem because also the handedness of the processes $B \to B-3$ and $B \to B+3$ is important. If $B \to B+3$ creates right-handed baryons, the process $B \to B-3$ would happen with the same frequency but would create left-handed antibaryons. As the baryon number does not distinguish between left-handed and right-handed baryons, and handedness is for massive particles not important as it can be changed by a Lorentz boost, the two processes which create three baryons and create three antibaryons again could cancel. Thus, in order to make the creation of baryons more probable not only C but also CP must be violated in the Standard Model. This is in fact the case, but the difference is that while C and P violation is maximal in the weak interactions, CP violation is only a very small effect.

Departure from Thermal Equilibrium

So far the thinking has been that there is one reaction going on. There is a state and the question is whether the baryon number is increased by three or decreased by three. However, the early universe was a hot mess where not single interactions happened at a time. In this soup of highly energetic particles a lot of interactions were going on and particles collided, appeared and disappeared. However, in a large system in thermal equilibrium one must expect that the rate of reaction in one direction equals the rate of reaction in the opposite direction. The reverse process of $B = 0 \to B + 3$ is $B + 3 \to B = 0$. If one starts with a symmetric universe with no net baryon number, it evolves with the process $B = 0 \to B + 3$ and might create baryons, but once there are enough baryons available they will start running the process as $B + 3 \to B = 0$ backwards.

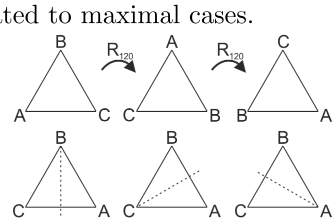
So it turns out that one needs a departure from equilibrium which will allow a dominance of reaction rates in one direction over the other. The early universe was not in an equilibrium state. The sphaleron process discussed here can be seen as a first-order phase transition which typically proceed via bubble nucleation (localized regions of the new phase). Bubbles with the new vacuum are formed and grow taking over old vacuum regions or collide with other bubbles as in boiling water. There are huge differences inside and outside of a bubble, and the bubble wall corresponds to a highly non-equilibrium configuration.

Hence all three Sakharov requirement can be satisfied by the Standard Model but while the pieces are there, the details are still to be completely worked out. Non-perturbative calculations are hard. There are cosmological effects to be taken into account. The problem of the baryonic asymmetry is still considered a problem in the cosmological aspects of Particle Physics which is not yet completely solved.

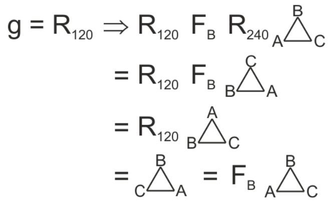
The Monstrous Moonshine Conjecture

Often in mathematics one can find arbitrarily complicated or large examples of structures one defines. For example in group theory, one can consider SO(N) and is free to take N to be as large a positive integer as one likes. Obviously if one asks what dimensions of spheres can be embedded in thirteen dimensions, this restricts the number of dimensions to be considered. What is really weird is when the definitions do not rely on numbers that are in any way or at least in an obvious way related to maximal cases.

As an example, the triangle group Tri is a discrete group with the elements $\{I, R_{120}, R_{240}, F_B, F_C, F_A\}$ and therefore with order $|\text{Tri}| = 6$. One can easily identify some subgroups such as $\{I, R_{120}, R_{240}\}$ or $\{I, F_A\}$.



It is often quite useful to identify so-called normal subgroups. A *normal subgroup* H of G is such that $gHg^{-1} = H$ for any $g \in G$. This does not mean $ghg^{-1} = h$ for all $h \in H$, but rather that gHg^{-1} takes the whole set of H and returns the whole set again.



For Tri the subgroup $\{I, F_B\}$ is a normal subgroup because $gIg^{-1} = I$ and $gF_Bg^{-1} = F_B$ for any $g \in$ Tri. The example $g = R_{120}$ with $g^{-1} = R_{240}$ gives $R_{120}F_BR_{240}I = F_BI$.

It is useful to define a *simple group* as a group G for which the only normal subgroups are $H = \{I\}$ and $H = G$ itself. An interesting question in mathematics was to try and classify all of the finite simple groups. An example is the cyclic group Z_n generated by g such that $g^n = I$. So the elements are the set $\{I, g, g^2, ..., g^{n-1}\}$. Clearly this can be generalized to arbitrarily large n. The order is determined by n because $|Z_n| = n$. All the cyclic groups form a family of groups.

In classifying all finite simple groups it was determined that there are 18 families of these, each family consisting of similar definitions but with one or more discrete parameters to adjust. For example Z_n is a finite simple group with one parameter n. In all cases the order is determined by the parameter choices, and the parameters can be arbitrarily high.

So far this is not that surprising. What is surprising is the existence of so-called "sporadic" groups. They are finite simple groups which do not belong to any of the 18 families, and which are "disconnected" in a parameter sense from each other. The smallest, and one of the earliest understood examples is called M_{11} and is the group of the Mathieu sharp 4-transitive group acting on eleven symbols. Its order is merely $|M_{11}| = 7920$. The largest of the sporadic groups is called monster group $\mathbb{M}$ with $|\mathbb{M}| \approx 8.08 \cdot 10^{54}$. The next largest group is baby-monster with a mere 10^{33} elements. The question is why such a large and isolated mathematical structure exists. This is not completely understood.

The two smallest complex representations of $\mathbb{M}$ other than the identity with dimension 1 have dimensions 196 883 and 21 296 876 respectively.

If one takes the upper half of the complex plane $\mathrm{Im}(z) > 0$ and maps it to the Riemann sphere $\tilde{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ using the function $j(z)$, to consider something completely unrelated as it seems, then the function $j(z)$ is elliptic, modular and has a Laurent expansion which is $j(z) = \frac{1}{q} + 744 + 196\,884q + 21\,493\,760q^2 + ...$ in terms of $q = \exp(2\pi iz)$. The coefficients have the quite surprising properties $196\,884 = 1 + 196\,883$ and $21\,493\,760 = 1 + 196\,883 + 21\,296\,876$.

Based on this observation a conjecture was made that these two different constructions must somehow be related. But most considered the idea so far fetched that they called it moonshine. Hence the conjecture is called monstrous moonshine. The final understanding of the connection came from a surprising and completely unrelated side, and the monstrous moonshine conjecture was eventually proven through the vertex operator algebras that appear in String Theory.

When one does Quantum Field Theory with zero-dimensional particles as excitations one can do it in any dimensions with any gauge groups SU(N), SO(N) where N is any integer $1, ..., \infty$. However, in String Theory the only change one makes compared to Quantum Field Theory is going from zero to one. Zero-dimensional point particles become one-dimensional strings. The change looks simple but has strong consequences. String Theory does not work in any dimensions, and one can also not just pick any gauge group one wants. The dimensions are 10, 11 or 26, and the gauge groups with specific numbers of generators become preferred. String Theory is what it is and one cannot change things arbitrarily. It is in a certain sense complete and selfcontained as a set of ideas and waiting out there like the monster group, but it is far from being completely understood.