# Particle Physics

Rainer F. Hauser
rainer.hauser@gmail.com

March 25, 2020

**Abstract**

Particle Physics is a very active and attractive field of physics. Leonard Susskind from the Stanford University held lectures in 2009 and 2010 covering this topic. His lectures were available on YouTube at the time this transcript has been assembled and may as "Particle Physics 1: Basic Concepts", "Particle Physics 2: Standard Model", "Particle Physics 3: Supersymmetry and Grand Unification" still be available today. The Higgs-boson had not been detected at that time, but plays an important role in these lectures. As usual for the courses by Susskind, minimal theoretical apparatus is used, but good knowledge in Classical Mechanics, Special Relativity and Quantum Mechanics is useful.

## 1 Basic Concepts

### 1.1 Matter or Field

The question whether matter is continuous or discrete is old. The believe today is that both views are correct because fields with waves are continuous and their quanta forming particles are discrete, but both views are also wrong due to Quantum Mechanics. Dalton detected that matter comes in units leading to the periodic table of the elements in chemistry, Becquerel found radioactivity with the alpha, beta and gamma decay, and the development of Quantum Mechanics brought the particle-wave-duality.

Gamma radiation is light, Maxwell's equations led to the conclusion that electromagnetic waves including light are fields, and Einstein demonstrated with his interpretation of the photoelectric effect that light consists of particles called photons. The famous two-slit experiment shows that not only light but also electrons and other particles belonging to matter produce interference patterns.

### 1.2 Properties of Waves and Particles

A wave has a wavelength $\lambda$ and a period $T$. The speed of a wave is $\lambda/T$ which is denoted by $c$ for light. The number of cycles per unit time is the frequency $f = 1/T$, and, preferred by physicists, the number of radians per second is the angular frequency $\omega = 2\pi f$ which is $\omega = 2\pi c/\lambda$ for light. The electromagnetic waves range from radio waves with arbitrarily long wavelength via microwaves, infrared, visible light, ultraviolet finally to X rays and gamma rays (from radioactivity) with very short wavelength.

For a simple wave of the form $e^{i(kx-\omega t)}$, there is one single velocity $v = \lambda/T = \omega/k$ called the phase velocity $v_\mathrm{p}$ with the wave number $k = 2\pi/\lambda$. Wave packages where $\omega(k)$ depends on $k$ have in addition to the phase velocity a group velocity $v_\mathrm{g} = d\omega/dk$ which is the relevant velocity of the package.

Particle Physics is basically Relativistic Quantum Mechanics where massive particles have an energy $E = mc^2$ coming from their mass. The mass $m$ does not depend on the velocity $v$, but the energy $E$ does because $E = mc^2$ is the energy at rest, and a particle has additional energy in form of kinetic energy when in motion. Photons have no mass, but they have energy which comes in units $E_p = \hbar\omega = hf$ where $h = \hbar \cdot 2\pi$ is Planck's constant. The energy of a ray is a multiple $E_\mathrm{ray} = n\hbar\omega$ with $n \in \mathbb{N}$, and the amplitude $A$ of the ray is proportional to $\sqrt{n}$ because $E_\mathrm{ray}$ is proportional to $A^2$.

The official units in physics are meter, second and kilogram, but different areas in physics use other units for simplicity. In astrophysics, for example, distances are measured in lightyears and time in years. The speed of light is therefore $c = 1$ in these units. In Particle Physics, on the other hand, speed of light and Planck's constant are often set $c = 1$ and $\hbar = 1$. Speed of light $c$ and Planck's constant $h$ or $\hbar$ together with the gravitational constant $G$ constrain all objects, because $c$ is the maximal speed, the uncertainty principle contains $\hbar$, and $G$ determines the attraction between objects. In conventional units

$$c = 2.99762458 \cdot 10^8 \, \frac{\mathrm{m}}{\mathrm{s}} \qquad \hbar = 1.0545571628 \cdot 10^{-34} \, \frac{\mathrm{kg\,m}^2}{\mathrm{s}}$$

are the speed of light and Planck's constant. The gravitational constant $G$ can practically be ignored in Particle Physics with the current experimental limitations today.

Besides energy also momentum is an important quantity since both are conserved. A particle at rest has no momentum because momentum is defined non-relativistically as $\vec{p} = m\vec{v}$. This definition cannot be used relativistically because photons have no mass but still have momentum. If light is very collimated, the momentum points in the direction of the ray, and its magnitude is $|p| = E/c$. The momentum of a photon is therefore $p = \hbar\omega/c$ or $p = h/\lambda$ using $E = \hbar\omega$, $\omega\lambda = 2\pi\, c$ and $h = 2\pi\,\hbar$. In other words, the shorter the wavelength is, the bigger is the momentum. If one wants to see very small objects, one needs a very small wavelength, and that leads to a very strong acceleration of the observed object. There are limits. If the Planck length is reached, physics starts to behave differently, but to get the necessary momentum, one would need a linear accelerator in the order of the size of the galaxy.

To summarize, the equations

$$E = hf = \hbar\omega \qquad p = \frac{h}{\lambda} \tag{1.1}$$

hold for any particle using the definition $\omega = 2\pi\, f$. (The direction of the momentum $p$ is the direction of the motion.) For particles with the speed of light $c$ such as photons also the equation $\lambda f = c$ is valid. This is a good approximation for neutrinos, but these particles can also slow down below the speed of light. For a slowly moving particle such as an electron with non-relativistic speed, the relation between energy $E$ and momentum $p$ is

$$v = \frac{p}{m} \qquad E = \frac{1}{2}mv^2 = \frac{1}{2}\frac{p^2}{m} \qquad \Rightarrow \qquad f = \frac{1}{2m}\frac{h}{\lambda^2}$$

leading to a less famous version of the Schrödinger equation based on (1.1) where $f$ is the time-dependent and $h/(2m\lambda^2)$ the space-dependent part.

## 1.3 Harmonic Oscillator

Computers cannot handle infinity. Thus, computer simulations of a wave on a one-dimensional string of length $D$ bounce back and forth such that the boundary conditions result in a violation of conservation of momentum. As a solution, one can connect beginning and end of the string to produce a loop. The wave becomes periodic, and momentum is conserved, but momentum is now quantized such that $\lambda = D/n$ with $n \in \mathbb{N}$. Otherwise, the wave would not match because fields are continuous and make no jumps. Thus, $p = h/\lambda = nh/D$ and momentum is quantized in multiples of $h/D$. If $D$ is large, one gets a dense distribution of possible momenta. If the loop is a circle with radius $R$, $p = nh/D = nh/2\pi R = n\hbar/R$, and angular momentum is $L = n\hbar$ since $L = mvR = pR$ where $L = pR$ is generally valid while $L = mvR$ is only valid non-relativistically.

Systems in an equilibrium state start oscillating around this equilibrium state when disturbed, and these oscillations result in waves. The harmonic oscillator has a frequency and is used to model such oscillations. An example is light that can be separated into different wavelength and amplitudes. Classically, the minimum energy of a harmonic oscillator is zero. Due to the uncertainty principle, this is not so in Quantum Mechanics, but one can define the lowest energy as 0 because only differences of energies play a role in physics. Thus, $E = n\hbar\omega$ for $n \in \mathbb{N}$ are the energy eigenvalues, and the eigenvector as the corresponding state can be denoted by the ket-vector $|n\rangle$. This energy of $n$ quanta is the same as the energy of $n$ photons with frequency $\omega$.

An operator $a^+$ for raising and an operator $a^-$ for lowering the eigenvectors can be defined by

$$a^+ |n\rangle = \sqrt{n+1}\, |n+1\rangle \qquad\qquad a^- |n\rangle = \sqrt{n}\, |n-1\rangle \tag{1.2}$$

as known from introductions to Quantum Mechanics. The lowering operator returns 0 when applied to the ground-state. Because $a^+a^- |n\rangle = a^+\sqrt{n}\, |n-1\rangle = \sqrt{n}a^+ |n-1\rangle = \sqrt{n}\sqrt{(n-1)+1}\, |n\rangle = n\, |n\rangle$, $n$ is the eigenvalue to the eigenvector $|n\rangle$ of the operator $a^+a^-$. The energy of the harmonic oscillator can be written as $h\omega\, a^+a^- = nh\omega$. The commutator $[a^-, a^+]$ of the raising and lowering operator is not zero because $a^-a^+ |n\rangle = \sqrt{n+1}a^- |n+1\rangle = n+1\, |n\rangle$ and $a^+a^- |n\rangle = n\, |n\rangle$ but $[a^-, a^+] = 1$. One can therefore not measure $a^+$ and $a^-$ simultaneously. They raise and lower the energy by one quantum, respectively. If the oscillation corresponds to an electromagnetic oscillation, they add and subtract photons of the corresponding frequency. Therefore, they are also called creation and annihilation operators.

## 1.4 Quantum Fields

In the loop with length $D$, each frequency corresponds to a harmonic oscillator. The state can be denoted by $|n_0\, n_1\, n_2 \ldots\rangle$ where $n_j$ is the number of quanta with energy $jh\omega$. This corresponds to the harmonics, for example, of a violin string. A wave $e^{ikx}$ on the loop with length $D$ must fulfill $e^{ikD} = e^0 = 1$, and $kD$ must be a multiple of $2\pi$. The wavelength $\lambda$ is quantized through $k\lambda = 2\pi$, and the momentum $p = 2\pi\hbar/\lambda$ is quantized as well through $p = \hbar k$. The quantity $k$ is called the wave number. Values $k < 0$ correspond to waves going left, and values $k > 0$ to waves going right. The occupation number $n(k)$ representing the number of particles (or quanta) with wave number $k$ can be used to specify the state of a system as $|\ldots n(1)\, n(2)\, n(3) \ldots\rangle$. The change of the particle number can be described using creation and annihilation operators.

Experiments in Particle Physics deal with creation and annihilation of particles. Quantum field theory can therefore also be described using similar creation and annihilation operators as the ones for the harmonic oscillator. The field $\Psi(x)$ can be modeled classically as a sum over some values $\alpha(k)$ times the wave corresponding to the wave number $k$. The real and the imaginary part of $\alpha(k)$ are observables. If one replaces $\alpha(k)$ by $a^-(k)$ and $\alpha^*(k)$ by $a^+(k)$ from (1.2), the field becomes quantized and the quantum field becomes the two operators

$$\Psi(x) = \sum_k a^-(k)e^{ikx} \qquad\qquad \Psi^\dagger(x) = \sum_k a^+(k)e^{-ikx} \tag{1.3}$$

which are obviously related to the Fourier transform in classical physics. The vacuum which is the empty field and which is also the ground-state is denoted by the state $|\ldots 0\,0\,0 \ldots\rangle$ or simply by $|0\rangle$.

The operator $a^+(k)$ creates a particle with momentum $\hbar k$ (or $k$ when $\hbar$ is set to 1). The situation that a particle scatters by colliding with something such that the momentum changes can be modeled using $a^+(5)a^-(3)\, |\ldots 0\,0\,1\,0\,0\,0 \ldots\rangle = a^+(5)\, |\ldots 0\,0\,0\,0\,0\,0 \ldots\rangle = |\ldots 0\,0\,0\,0\,1\,0 \ldots\rangle$, for example, using (1.2).

The result of the operator $\Psi^\dagger(x)$ in the right equation of (1.3) is a superposition of states with different momenta $k$ (with $\hbar = 1$) all with the same probability but with a definite position $x$. It is therefore the operator for the creation of a particle at position $x$. Similarly, $\Psi(x)$ annihilates a particle at position $x$. The state $a^+(k)\, |\ldots\rangle$ is a state in momentum space, $\Psi^\dagger(x)\, |\ldots\rangle$ is a state in position space, and $a^-(k)\, |\ldots\rangle$ and $\Psi(x)\, |\ldots\rangle$ are similarly states in momentum space and in position space, respectively.

When a particle somehow decays into two particles at position $x$, then it gets annihilated at position $x$ and the two emitted particles are created also at position $x$ (or at least very close to $x$). This is the property of locality in field theory. The situation may occur when a photon comes in, hits an atom and two photons go off. It can be modeled as $\Psi^\dagger(x)\Psi^\dagger(x)\Psi(x)\, |\ldots\rangle$. Assuming that the incoming photon has momentum $k_i$ then the initial state is $|\ldots 0 \ldots 0\,1\,0 \ldots\rangle$ where the occupation number $n(k_i) = 1$ and all the others are $n(k) = 0$. The annihilation can be modeled by

$$\sum_k a^-(k)e^{+ikx}\, |\ldots 0 \ldots 0\,1\,0 \ldots\rangle = a^-(k_i)e^{+ik_ix}\, |\ldots 0 \ldots 0\,1\,0 \ldots\rangle = e^{+ik_ix}\, |\ldots 0 \ldots 0\,0\,0 \ldots\rangle$$

because only the operator $a^-(k_i)$ contributes in the sum and its effect is to reduce $n(k_i)$ by one.

The whole process with the incoming photon and the two outgoing photons becomes

$$\sum_m a^+(m)e^{-imx} \sum_l a^+(l)e^{-ilx} \sum_k a^-(k)e^{+ikx} |k_i\rangle = \sum_m a^+(m)e^{-imx} \sum_l a^+(l)e^{-ilx}\, e^{+ik_ix} |0\rangle =$$
$$\sum_{l,m} e^{-imx}\, e^{-ilx}\, e^{+ik_ix} |l\, m\rangle = \sum_{l,m} e^{i(k_i-l-m)x} |l\, m\rangle$$

where – changing notation – the states $|k_i\rangle$ and $|l\, m\rangle$ describe here the state with only one particle with momentum $k_i$ respectively the state of two particles one with momentum $l$ and one with momentum $m$.

## 1.5 Bosons

The above equations in (1.2) and (1.3) are only valid for particles which are bosons. If there is already one particle with $k$, the probability for $a^+(k)\,|1\rangle = \sqrt{2}\,|2\rangle$ is twice as big compared to the situation where there is no particle with $k$ where the probability is $a^+(k)\,|0\rangle = \sqrt{1}\,|1\rangle$. Thus, bosons prefer to be in a state together with others. This is called stimulated emission, while spontaneous emissions happen when no other particle with the same $k$ is available. In other words, if there are already photons (as an example of a kind of bosons) with a certain $k$, the probability is higher for the creation of another photon with the same $k$ than for a photon with a value $k$ not already present. Only bosons have this tendency but fermions do not because of the Pauli exclusion principle.

The field $\Psi(x) = \sum a^-(k)e^{ikx}$ together with $[a^-(k), a^+(l)] = \delta_{kl}$, and $[a^+(k), a^+(l)] = [a^-(k), a^-(l)] = 0$ as the commutator relations is the simplest Quantum Field Theory with the assumption that there is only one kind of particles. On the loop with length $D$,

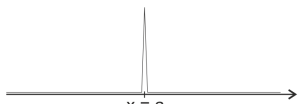
$$\int \Psi^\dagger(x)\, \Psi(x)dx = \sum_{kl} a^+(k)\, a^-(l) \int e^{i(l-k)x}dx = D \cdot \sum_k a^+(k)a^-(k) = N \cdot D$$

because of (1.5), and because $a^+(k)a^-(k)$ is the occupation number for $k$ such that $N$ is the total number of particles. Thus,

$$\frac{1}{D}\int \Psi^\dagger(x)\Psi(x)dx = N \qquad\qquad \Psi^\dagger(x)\Psi(x) = \rho(x) \tag{1.4}$$

give the total number $N$ of particles and (apart from a factor $D$) the density $\rho(x)$ at position $x$, respectively. For large numbers $N$, $\Psi$ behaves like a classical field. (This is not completely correct because the relativistic behavior of photons has not been taken into account yet.)

The above integral is based on the Dirac delta-function $\delta(x - a)$ which is actually not a real function but a distribution and which is a continuous analogon to the Kronecker delta $\delta_{kl}$. It is everywhere zero except for $x = a$ such that $\int \delta(x - a)dx = 1$. On the loop with length $D$ with $k = 2\pi n/D$




$$\int_{-\frac{D}{2}}^{+\frac{D}{2}} e^{ikx}dx = \begin{cases} D & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases} \qquad \int_{-\infty}^{+\infty} e^{ikx}dx = 2\pi\, \delta(k) \qquad \int_{-\infty}^{+\infty} e^{ikx}dk = 2\pi\, \delta(x) \tag{1.5}$$

for $D \to \infty$. The delta function plays an important role when either the position is exactly known and momentum is completely unknown or vice versa.

## 1.6 More on Quantum Fields

A ket is a symbolic notation for a quantum state. It is half of a bra(c)ket, and the other half is a bra. They are just notational devices. The initial state of a system will be defined as a ket-vector $|\text{initial}\rangle$ and the final state $\langle\text{final}|$ as a bra-vector. Putting a bra-vector next to a ket-vector is a symbolic manipulation making a full bracket. In the end what comes out of symbolic manipulations are numbers. The inner product $\langle s_1|s_2\rangle$ of two state vectors $s_1$ and $s_2$ is a number, and this means $\langle m|n\rangle = \delta_{mn}$ for the vectors $|m\rangle$ and $|n\rangle$ of an orthonormal basis as in the case of the harmonic oscillator above.

The two formulas (1.2) show how the raising and lowering operators $a^+$ and $a^-$ of the harmonic oscillator act on ket-vectors, and the question is how they act on bra-vectors. From $\langle n|a^+|m\rangle = \sqrt{m+1}\,\langle m+1|n\rangle$ follows that the result is zero except for $m+1=n$. Thus, acting on a bra-vector the operator $a^+$ must decrease the index by one, and the formulas are

$$\langle n|\,a^+ = \langle n-1|\,\sqrt{n} \qquad\qquad \langle n|\,a^- = \langle n+1|\,\sqrt{n} \tag{1.6}$$

for the bra-vectors. The expectation value of $a^+a^-$ is therefore $\langle n|a^+a^-|n\rangle = n\,\langle n|n\rangle = n$.

The only possibilities for gaining knowledge in physics of this scale are scattering experiments. Therefore the main operators for processes in the simplest quantum field $\Psi(x)$ introduced above and its hermitian conjugate $\Psi^\dagger(x)$ are

$$\Psi^\dagger(x,t) = \sum_k a^+(k)e^{-ikx}e^{i\omega(k)t} \qquad\qquad \Psi(x,t) = \sum_k a^-(k)e^{ikx}e^{-i\omega(k)t} \tag{1.7}$$

but where $x$ and $k$ are assumed to be three-dimensional (and could be written as $\vec{x}$ and $\vec{k}$). The time-dependence added to (1.3) is needed because real fields oscillate.

For a particle with speed much less than the speed of light, energy expressed in momentum is $E = p^2/2m$ and therefore $\omega(k) = k^2/2m$ because $E = \omega$ and $p = k$ (with $\hbar = 1$). Inserted into (1.7) gives

$$\frac{\partial}{\partial t}\Psi^\dagger(x,t) = \sum_k i\omega(k)a^+(k)e^{-ikx}e^{i\omega(k)t} \qquad\qquad \frac{\partial^2}{\partial x^2}\Psi^\dagger(x,t) = \sum_k (-k^2)a^+(k)e^{-ikx}e^{i\omega(k)t}$$

as the wave equation from which the Schrödinger equation

$$i\frac{\partial}{\partial t}\Psi(x) = \frac{1}{2m}\frac{\partial^2}{\partial x^2}\Psi(x) \tag{1.8}$$

follows because $k^2 = \omega(k)\,2m$. The difference to the Schrödinger equation in Quantum Mechanics is that here $\Psi(x,t)$ is not a wave function but an operator, and this operator depends on the kind of particle. (Note however that this quantum field theory defined by $\Psi^\dagger(x,t)$ and $\Psi(x,t)$ in (1.7) is a simplified model and does not describe any real particles.)

This equation can be used to describe processes in which a particle scatters of a very heavy target that does not recoil. The momentum of the particle changes but energy is conserved. One can think of the process as a particle with momentum $k_\mathrm{i}$ being annihilated when it hits the target and a new particle with momentum $k_\mathrm{f}$ being created. (Initially, only one particle exists that has momentum $k_\mathrm{i}$, and after the scattering only one particle exists that has momentum $k_\mathrm{f}$.) The target is located at $x = 0$, and the process can happen at any time. Thus, the process itself and the probability averaged over time is

$$\langle k_\mathrm{f}|\int_t dt\,\Psi^\dagger(0,t)\Psi(0,t)\,|k_\mathrm{i}\rangle \qquad\qquad \left|\langle k_\mathrm{f}|\,g\int_t dt\,\Psi^\dagger(0,t)\Psi(0,t)\,|k_\mathrm{i}\rangle\right|^2$$

with a coupling constant $g$. The strength of the scattering is described by the coupling constant which comes from experiments. The scattering of a meson on a proton, for example, is much stronger that the scattering of a photon on a proton. The coupling constant of the meson-scattering is therefore much bigger than the coupling constant of the photon-scattering, and the coupling constant for scattering a neutrino on a proton is even smaller.

The probability of the above scattering process can be calculated by using the definition of $\Psi$ and $\Psi^\dagger$ in terms of the creation and annihilation operators, but the calculation is only valid for very small values of $g$. Because $x = 0$, $\langle k_\mathrm{f}|\,a^+(k_\mathrm{f}) = \langle 0|$, $a^-(k_\mathrm{i})\,|k_\mathrm{i}\rangle = |0\rangle$ and $\langle 0|0\rangle = 1$, the probability amplitude becomes

$$\langle k_\mathrm{f}|\,g\int_t dt\,a^+(k_\mathrm{f})e^{i\omega(k_\mathrm{f})t}\,a^-(k_\mathrm{i})e^{-i\omega(k_\mathrm{i})t}|k_\mathrm{i}\rangle = \langle 0|\,g\int_t dt\,e^{i(\omega(k_\mathrm{f})-\omega(k_\mathrm{i}))t}|0\rangle = 2\pi\,g\,\delta(\omega(k_\mathrm{f})-\omega(k_\mathrm{i}))$$

using (1.5). The delta-function comes from the integral over time, and it is only non-zero if $\omega(k_\mathrm{f}) = \omega(k_\mathrm{i})$. Thus, there is a connection between the conservation of energy and the fact that there is no preferred time which corresponds to time-translation symmetry. The probability becomes $4\pi^2g^2$ if the initial energy is equal to the final energy. The momentum $k_\mathrm{f}$ can be anything as long as $\omega(k_\mathrm{f}) = \omega(k_\mathrm{i})$.

The scattering process is actually not as simple as shown here because the particle can come in, scatter and go out, but may come in again and scatter this way several times. There is an infinite sequence of possibilities where each possibility gets an additional factor of $g$. Thus, higher-order terms can only be ignored if $g$ is small compared to 1. Actually, coupling constants $g$ can be complex numbers when time-reversal invariance is violated.

The position of the target has been set to $x = 0$, but a different coordinate would not change the result, because it would just add another factor $e^{i(k_\mathrm{i}-k_\mathrm{f})x_\mathrm{T}}$ where $x_\mathrm{T}$ is the location of the target. This factor becomes 1 when multiplied with the complex conjugate for the probability.

If the scattering must happen at a specific location but can happen at any time, the sum of the frequencies of all incoming particles must be equal to the sum of the frequencies of all outgoing particles. This is guaranteed by the resulting delta function. Therefore, time-translation invariance leads to conservation of energy. Similarly, space-translation invariance leads to conservation of momentum as the example of a particle decaying into two particles at time $t = 0$ indicates. The process can be modeled as

$$\langle k_1\, k_2| \int_x dx\, \Psi^\dagger(x,0)\Psi^\dagger(x,0)\Psi(x,0)\, |k_\mathrm{i}\rangle$$

leading to a delta-function due to the integral over $e^{i(k_\mathrm{i}-k_1-k_2)x}$ and therefore to $k_\mathrm{i} = k_1 + k_2$.

Each kind of particle has its own separate field. Thus, electrons, for example, have a different field than photons. If $\Psi$ describes the electron-field in the above scattering situation (assuming a bosonic kind of electrons for the argument) the electric charge would not have changed. If one, however, imagines a scattering process where one electron comes in and two electrons go out as modeled by $\Psi^\dagger(0,t)\Psi^\dagger(0,t)\Psi(0,t)$, nature would not allow this process because it violates electric charge conservation. The rule is obviously that the number of incoming electrons must be the same as the number of outgoing electrons, and this rule can be written mathematically: Electric charge is preserved if and only if the equation does not change under the transformation $\Psi \to e^{i\alpha}\Psi$, $\Psi^\dagger \to e^{-i\alpha}\Psi^\dagger$ or, in words, if the equation does not change due to multiplication with a phase.

There are three different conservation laws so far. Energy conservation means time-translation invariance, momentum conservation means space-translation invariance, and electric charge conservation means invariance under multiplication with a phase. Energy conservation and momentum conservation lead to very similar delta-functions, energy conservation through an integral over time and momentum conservation through an integral over space.

## 1.7 More on Waves

Only the group velocity of a wave package has a physical meaning but not the phase velocity. For a plain sine-wave $\sin(kx - \omega t)$, the point $x = t = 0$ moves such that $kx - \omega t = 0$ or $x = (\omega/k)t$. Thus, $\omega/k$ is the velocity of a peek of the sine-wave and is called the phase velocity. Because of $\omega = 2\pi/T$ and $k = 2\pi/\lambda$ the phase velocity is also $\lambda/T$.

Schrödinger fields have energy and momentum $E = \omega$ and $p = k$ (with $\hbar$ set to 1) and satisfy therefore $\omega = k^2/2m$ such that the phase velocity is $v_\mathrm{p} = k/2m$. The velocity of a particle is $p/m = k/m$, and this is not the phase velocity. Adding a constant $a$ to $\omega$ such that $\omega = k^2/2m + a$ changes the phase velocity to $\omega/k = k/2m + a/k$, but this has no physical significance because only differences of energy have a meaning. The constant $a$ leads to a phase $e^{-iat}$ in the field operator $\Psi(x,t)$ which cancels when multiplied with the complex conjugate $\Psi^\dagger(x,t)$. All of the quantities of physical interest that are made out of the Schrödinger field involve $\Psi$ times its complex conjugate where the factor $e^{-iat}$ cancels out.

Several waves can add up to a wave packages due to constructive and destructive interference. If one adds two waves such as $\sin(k_1 x - \omega(k_1)t) + \sin(k_2 x - \omega(k_2)t)$ with very close momenta $k_1$ and $k_2$, the constructive interference are strongest when $k_1 x - \omega_1 t = k_2 x - \omega_2 t$ with $\omega_i = \omega(k_i)$. Written in the form $x = (\omega_1 - \omega_2)t/(k_1 - k_2)$ with $k_1$ and $k_2$ infinitesimally close this becomes the derivative $d\omega/dk$. Because of $x = (d\omega/dk)t$ this is also a velocity and is called the group velocity $v_\mathrm{g}$ which is as a derivative independent of an added constant $a$ and which is exactly the velocity of a non-relativistic particle.

For a relativistic wave, the relation between energy $E$ and momentum $p$ is $E = \sqrt{p^2 + m^2}$ (with the usual conventions $c = \hbar = 1$) because the square of the energy-momentum 4-vector must be the same for a frame in motion $-E^2 + p^2$ and a frame at rest $-m^2$. Thus, $\omega = \sqrt{k^2 + m^2}$ for any particle. For a massless particle this equation reduces to $E = |p|$ or $\omega = |k|$. This means that the phase velocity $v_p$ and the group velocity $v_g$ are equal and become

$$v_p = v_g = 1$$

for $m = 0$. This is the speed of light. These two velocities are

$$v_p = \frac{\omega}{k} = \sqrt{\frac{k^2 + m^2}{k^2}} \qquad\qquad v_g = \frac{d\omega}{dk} = \frac{k}{\sqrt{k^2 + m^2}} = \sqrt{\frac{k^2}{k^2 + m^2}}$$

for $m \neq 0$. The phase velocity $v_p$ is greater than the speed of light and the group velocity smaller.

## 1.8 Fermions

There is no limit to the number of bosons with the same state, and they have the tendency to have the same momentum as the majority of the other bosons. Two fermions in contrast cannot be in the same state because of the Pauli exclusion principle. Therefore, fermions can never behave classically, because it is not possible to put large numbers of fermions into the same state as it is possible for bosons.

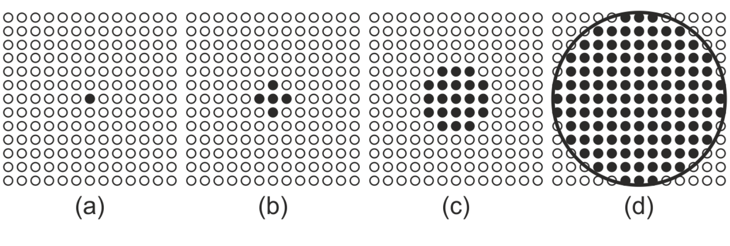
Non-relativistic fermion fields are defined the same way as boson fields through (1.3), but the operators $\Psi(x)$ and $\Psi^\dagger(x)$ cannot violate the Pauli exclusion principle. The creation and annihilation operators for fermions are completely defined by

$$c^+ |0\rangle = |1\rangle \qquad c^+ |1\rangle = 0 \qquad c^- |0\rangle = 0 \qquad c^- |1\rangle = |0\rangle \tag{1.9}$$

and they therefore behave differently than the ones for bosons described by (1.2). To distinguish them from the bosonic operators $a^+$ and $a^-$ they have been called $c^+$ and $c^-$.

With $c^+c^- |0\rangle = 0$, $c^+c^- |1\rangle = |1\rangle$, $c^-c^+ |0\rangle = |0\rangle$, $c^-c^+ |1\rangle = 0$ the algebra is $(c^+c^- + c^-c^+) |0\rangle = |0\rangle$ and $(c^+c^- + c^-c^+) |1\rangle = |1\rangle$ and can be written as $\{c^+, c^-\} = \mathbf{I}$ with the anticommutator. (Note that this is analogous to the commutator $[a^+, a^-] = \mathbf{I}$ for bosons. In general, commutators for bosons become anticommutators for fermions.) The anticommutators $\{c^+, c^+\} = \{c^-, c^-\} = 0$ complete the algebra for the creation and annihilation operators of fermions. Since the creation and annihilation operators exist for different momenta $k$, the corresponding anticommutators are $\{c^+(k_1), c^-(k_2)\} = \delta_{k_1 k_2}$ and $\{c^+(k_1), c^+(k_2)\} = \{c^-(k_1), c^-(k_2)\} = 0$. Thus, two fermions of the same kind cannot be in the same momentum state. Because $\Psi^\dagger(0)\Psi^\dagger(0) |0\rangle = \sum_{k_1} \sum_{k_2} c^+(k_1)c^+(k_2) |0\rangle$ they can also not be in the same position state.

A system is a collection of degrees of freedom. A configuration of a system in classical physics corresponds to a state of the system in Quantum Mechanics. A fermion in state $(1/\sqrt{2})(|s_1\rangle + |s_2\rangle)$ is equally likely to be in state $|s_1\rangle$ and in state $|s_2\rangle$. Another ferminon of the same kind cannot be in this state as well. The two fermions, however, can be one in state $|s_1\rangle$ and one in state $|s_2\rangle$. The difference between bosons and fermions can be shown graphically. Momentum $k = (k_x, k_y, k_z)$ is modeled as a discrete two-dimensional space $(k_x, k_y)$ with the ground state in the middle. Bosons of the same kind can all be put into the same ground state with $k = 0$ as shown in (a) in the figure below. This is called a Bose condensate.




Fermions behave differently. The first fermion can go into the lowest state as in (a), but the next fermions of the same type have to find another state with a bit higher energy as in (b). Further fermions of this type fill a sphere in momentum space as in (c) and in (d). This sphere due to $E = k^2/2m$ is called

Fermi sphere. This needs a lot more energy than bosons, and the more fermions are there, the bigger is the energy of the system and the total momentum. With $N$ bosons in the ground state the next higher energy level is one boson in the first excited state and all the $N-1$ other bosons in the ground state.

If one moves one electron as an example of a fermion from just below the Fermi sphere to slightly outside the Fermi sphere, that does not change the total charge and changes the energy only a bit, but it leaves a hole in the Fermi sphere. The hole is an absence of negative charge but can be thought of as the presence of a positive charge. The energy needed to move the electron out is the energy to move the electron from its original place to the Fermi sphere and from there to the final place. The first part of positive energy can be seen as the energy of the hole, and the other part as the additional energy of the electron. The electron can move around outside of the Fermi sphere and then all of a sudden drop back into the hole. One or more photons are emitted to compensate for the reduced energy, and the hole is annihilated. The hole can be interpreted as a particle called positron.

In atomic physics an atom with as many electrons as protons, where all the electrons are in the lowest possible state, is the analog to the Fermi sphere. A photon comes along and kicks one electron into an excited state. It is a low-energy photon that cannot remove and therefore be absorbed by an electron from deep inside the Fermi sphere, but only an electron close to its surface. This process can be interpreted as a photon splitting into an electron (the excited electron) and a positron (the hole).

## 1.9 Dirac Equation Simplified

The very simple differential equation $\partial\Psi/\partial t = -\partial\Psi/\partial x$ has the solution $\Psi = e^{i(kx-\omega t)}$ representing a wave or particle moving in one direction with the speed of light because of $\omega = k$. The particle is assumed to be an electrically charged fermion. The value $k$ can be positive or negative and therefore so can $\omega$. This is a problem because electrons cannot have arbitrarily low energy as the ground state would become minus infinity. To solve this problem, Dirac assumed that all negative energy states are occupied, and this negative energy sea is the vacuum (the empty space). This assumption only works for fermions, because bosons cannot fill such a sea. When an electron with negative energy is moved out of this sea by the kick of a photon the electron gets a positive energy and a hole remains. Such a hole in this sea is a positive charge (a missing negative charge). It is called positron and is therefore an antiparticle.

Several things are wrong with this electron. It moves with the speed of light, and it moves only in one direction. With $v_\mathrm{g} = \omega/k$ it moves to the right, when $\omega$ and $k$ both are positive or negative, and it moves to the left, when $\omega$ and $k$ have different signs.

The electrons moving to the right and therefore satisfy $\partial\Psi_\mathrm{R}/\partial t + \partial\Psi_\mathrm{R}/\partial x = 0$ (with $\omega/k = +1$) fill the Dirac sea with a lot of negative momentum. A completely different field fulfilling $\partial\Psi_\mathrm{L}/\partial t - \partial\Psi_\mathrm{L}/\partial x = 0$ (with $\omega/k = -1$) gives a second kind of electrons that move to the left. If the Dirac sea is filled with the same number of right and left moving electrons, there is no longer negative momentum in the sea. So far, these particles and antiparticles are massless because they move with the speed of light. With

$$\Psi = \begin{pmatrix}\Psi_\mathrm{R}\\ \Psi_\mathrm{L}\end{pmatrix} \qquad \alpha = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \qquad \frac{\partial}{\partial t}\Psi = -\alpha\frac{\partial}{\partial x}\Psi \qquad \Rightarrow \qquad \begin{pmatrix}\frac{\partial}{\partial t}\Psi_\mathrm{R}\\ \frac{\partial}{\partial t}\Psi_\mathrm{L}\end{pmatrix} = -\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\frac{\partial}{\partial x}\Psi_\mathrm{R}\\ \frac{\partial}{\partial x}\Psi_\mathrm{L}\end{pmatrix}$$

the derivative leads to the equation $-i\omega = -\alpha(ik)$ or $\omega = \alpha\cdot k$, and this equation becomes $\omega = \alpha k + \beta m$ as the square root of the relativistic equation $\omega^2 = k^2 + m^2$ (or $\omega = \sqrt{k^2+m^2}$). It follows $\alpha\beta + \beta\alpha = 0$ because $\omega^2 = (\alpha k + \beta m)(\alpha k + \beta m) = \alpha^2 k^2 + \beta^2 m^2 + (\alpha\beta + \beta\alpha)km$. Thus, $\alpha$ and $\beta$ cannot be numbers. Further, $\alpha^2 = \beta^2 = \mathbf{I}$ and $\alpha \neq \beta$ is needed for $\omega^2 = k^2 + m^2$. The matrices and the resulting equation

$$\alpha = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \qquad \beta = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \qquad \Rightarrow \qquad i\begin{pmatrix}\frac{\partial}{\partial t}\Psi_\mathrm{R}\\ \frac{\partial}{\partial t}\Psi_\mathrm{L}\end{pmatrix} = -i\begin{pmatrix}\frac{\partial}{\partial x}\Psi_\mathrm{R}\\ -\frac{\partial}{\partial x}\Psi_\mathrm{L}\end{pmatrix} + m\begin{pmatrix}\Psi_\mathrm{L}\\ \Psi_\mathrm{R}\end{pmatrix}$$

fulfill these requirements. The result written as two separate equations is

$$i\frac{\partial}{\partial t}\Psi_\mathrm{R} = -i\frac{\partial}{\partial x}\Psi_\mathrm{R} + m\Psi_\mathrm{L} \qquad\qquad i\frac{\partial}{\partial t}\Psi_\mathrm{L} = +i\frac{\partial}{\partial x}\Psi_\mathrm{L} + m\Psi_\mathrm{R}$$

showing that the term for the mass mixes right and left moving particles. After introducing mass, $\Psi_\mathrm{R}$ and $\Psi_\mathrm{L}$ are therefore no longer separated right and left moving particles. These equations are Lorentz invariant because $\omega^2 = k^2 + m^2$ is Lorentz invariant.

The limit for electrons at rest (which is not possible for massless particles) is $\omega = m\beta$. This leads to

$$i\frac{\partial}{\partial t}\Psi_{\mathrm{R}} = m\Psi_{\mathrm{L}} \qquad i\frac{\partial}{\partial t}\Psi_{\mathrm{L}} = m\Psi_{\mathrm{R}} \qquad i\frac{\partial}{\partial t}\Psi_{+} = m\Psi_{+} \qquad i\frac{\partial}{\partial t}\Psi_{-} = -m\Psi_{-}$$

where $\Psi_{+} = (1/\sqrt{2})(\Psi_{\mathrm{R}} + \Psi_{\mathrm{L}})$ and $\Psi_{-} = (1/\sqrt{2})(\Psi_{\mathrm{R}} - \Psi_{\mathrm{L}})$ result from adding and subtracting $\Psi_{\mathrm{R}}$ and $\Psi_{\mathrm{L}}$, respectively, and where $\omega = m$ for $\Psi_{+}$ and $\omega = -m$ for $\Psi_{-}$. This shows that also particles at rest have positive and negative frequency and therefore positive and negative energy. The Dirac sea is filled with $\Psi_{-}$ corresponding to particles with negative energy. There are positive energy electrons and positive energy positrons corresponding to the holes in the Dirac sea, and both have positive mass.

The scalar Higgs field $\Phi(x)$ is a bosonic scalar field. It is connected to the Dirac equation by

$$i\frac{\partial}{\partial t}\Psi_{\mathrm{R}} = -i\frac{\partial}{\partial x}\Psi_{\mathrm{R}} + g\Phi\Psi_{\mathrm{L}} \qquad i\frac{\partial}{\partial t}\Psi_{\mathrm{L}} = i\frac{\partial}{\partial x}\Psi_{\mathrm{L}} + g\Phi\Psi_{\mathrm{R}}$$

where $g$ is a coupling constant. If the vacuum would be such that $\Phi = 0$ as one expects for empty space, the electron in empty space would be massless. However, if for some reason the energetics of the Higgs field favor the lowest energy having a non-zero value of $\Phi$, for example, a constant $\Phi \neq 0$, the quantity $g\Phi$ would play the role of the mass $m$.

Bosonic fields have potential energy, and this is energy $V(\Phi)$ which depends on the value of the field. This potential $V(\Phi)$ is symmetric and has its minimum not at $\Phi = 0$. Thus, there is more than one value of $\Phi$ corresponding to the state of lowest energy all giving rise to a mass term. The Higgs field can oscillate around the chosen state of lowest energy. The frequency with which the Higgs field vibrates is related to the mass of the Higgs particle, and the excitations of the Higgs field come in quanta which are the Higgs particle. If the Higgs field is coupled in an interesting way with the electron field, the Higgs field will react on collisions with the fermions with vibrations leading to Higgs particles. If the mass of the Higgs particle is large, it takes much energy to excite one single quantum.

## 1.10 Dirac Equation

The real Dirac equation in a world which has more directions than only left and right starts also from the equation $\omega = \alpha \cdot k + \beta \cdot m$ where $k$ is now a three-dimensional vector, and the condition for relativity is $\omega^2 = k_1^2 + k_2^2 + k_3^2 + m^2$. From $\omega^2 = (\alpha_1 k_1 + \alpha_2 k_2 + \alpha_3 k_3)(\alpha_1 k_1 + \alpha_2 k_2 + \alpha_3 k_3) = \omega^2 = k_1^2 + k_2^2 + k_3^2 + m^2$ follows $\alpha_1^2 = \alpha_2^2 = \alpha_3^2 = \beta^2 = \mathbf{I}$ and $\{\alpha_i, \alpha_j\} = 0$ for $i \neq j$ as well as $\{\alpha_i, \beta\} = 0$ for all $i$.

These conditions cannot be fulfilled by $2 \times 2$ matrices because there are only three Pauli matrices $\sigma_j$ (introduced below), but four matrices are needed. Also the $3 \times 3$ matrices do not provide a solution, but some $4 \times 4$ matrices do. Four matrices solving these conditions are

$$\alpha_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \alpha_2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \quad \alpha_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \quad \alpha_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}$$
$$\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad \beta = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & -\mathbf{I} \end{pmatrix} \quad \Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \end{pmatrix} \tag{1.10}$$

which are called Dirac matrices. They are not unique in a similar sense as orthogonal axes in space are not unique but all possible choices are equivalent. Note also that $\Psi$ has four components but is not a 4-vector, but a spinor. The Dirac equation (standing for four equations) becomes

$$i\frac{\partial}{\partial t}\Psi_p = -i(\alpha_j)_{pq}\frac{\partial}{\partial x^j}\Psi_q + \beta_{pq} m\Psi_q \qquad i\frac{\partial}{\partial t}\Psi_p = (\beta)_{pq} m\Psi_q \tag{1.11}$$

where the fermion on the left side has momentum and on the right side is at rest. Again, mass comes from mixing the different components of $\Psi$. Before Dirac wrote this equation, there was no idea of antiparticles, and spin was known but nobody knew where it comes from.

## 1.11 Spin

The spin is the angular momentum when the center of mass is at rest and ordinary momentum is therefore zero. Classically, an object can be set continuously into rotation, but in Quantum Mechanics spin is not continuous. The question is also whether one can bring an electron, for example, into rotation. It turned out that the smaller an object is, the more energy is needed to bring it into rotation. It takes very much energy to bring an electron into the first angular momentum and one does not see rotating electrons in the laboratory. It would probably become a different object. In Quantum Mechanics small objects have an amount of angular momentum which characterizes them and which is once and for all fixed. It is called the spin of the object.

Angular momentum is $L = r \times p = (yp_z - zp_y, zp_x - xp_z, xp_y - yp_x)$ with $r = (x, y, z) = (x_1, x_2, x_3)$ and $p = (p_x, p_y, p_z) = (p_1, p_2, p_3)$. It is a vector, has a length and a direction, and it depends on the origin of the coordinate system. Quantum Mechanics deals with operators and their commutation relations. The commutators needed here are $[x_j, p_k] = i\hbar\, \delta_{jk}$ and $[x_j, x_k] = [p_j, p_k] = 0$ such that $[L_x, L_y] = [yp_z - zp_y, zp_x - xp_z] = i\hbar L_z$ and similarly $[L_y, L_z] = i\hbar L_x$ and $[L_z, L_x] = i\hbar L_y$.

Adding and subtracting the components of $L$ such that $L_+ = L_x + iL_y$ and $L_- = L_x - iL_y$ are raising and lowering operators. The relevant commutators are $[L_+, L_z] = -L_+$ and $[L_-, L_z] = +L_-$ with $\hbar = 1$. One can only measure one direction of angular momentum, and $L_z \left|m\right\rangle = m \left|m\right\rangle$ is selected with its eigenvectors $\left|m\right\rangle$. $L_+ \left|m\right\rangle = (m + 1)L_+ \left|m\right\rangle$ because $L_z L_+ \left|m\right\rangle = (L_+ L_z + L_+) \left|m\right\rangle = L_+ L_z \left|m\right\rangle + L_+ \left|m\right\rangle$ and this means that $L_+ \left|m\right\rangle$ is an eigenvector of $L_z$ with eigenvalue $m + 1$. Similarly, $L_- \left|m\right\rangle$ is an eigenvector of $L_z$ with eigenvalue $m - 1$.

Because of $L_+ \left|m_{\max}\right\rangle = 0$ and $L_- \left|m_{\min}\right\rangle = 0$ the eigenvalues of $L_z$ are a finite set of discrete values between $m_{\min}$ and $m_{\max}$ in integer steps. There are only the two possibilities $m_{\min}, ..., -1, 0, +1, ..., m_{\max}$ and $m_{\min}, ..., -\frac{1}{2}, +\frac{1}{2}..., m_{\max}$ both with $m_{\min} = -m_{\max}$. The value $m_{\max}$ depends on the type of particle and is called its spin. Theory says that no elementary particle can have a spin greater than 2. Bosons have an integer spin. The Higgs boson has spin 0, the photon and the W- and Z-bosons have spin 1, and the graviton is assumed to have spin 2. Fermions have a half-integer spins. The electrons, quarks and neutrinos have spin $\frac{1}{2}$, and some hypothetical supersymmetric particles are assumed to have spin $\frac{3}{2}$.

The operator $L^2 = L_x^2 + L_y^2 + L_z^2$ is not $L_- L_+ + L_z^2$ because of the commutator of $L_x$ and $L_y$. It is thus $L^2 = L_- L_+ + L_z^2 + L_z$. Because $L^2 \left|m_{\max}\right\rangle = L_- L_+ \left|m_{\max}\right\rangle + (m_{\max}^2 + m_{\max}) \left|m_{\max}\right\rangle$, one eigenvalue of $L^2$ is $m_{\max}(m_{\max} + 1)$. (For large $m_{\max}$ this is approximately $m_{\max}^2$.) One can know $L^2$ and one of the components $L_i$ because $[L^2, L_i] = 0$. Consequently also $[L^2, L_+] = [L^2, L_-] = 0$, and therefore all eigenvectors of $L_-$ have the same eigenvalue for $L^2$ because of $L^2 L_- \left|m_{\max}\right\rangle = m_{\max}(m_{\max}+1)L_- \left|m_{\max}\right\rangle$.

The spin of an electron can be in one of the two states up and down denoted by $\left|u\right\rangle$ and $\left|d\right\rangle$. A general quantum state is $\alpha \left|u\right\rangle + \beta \left|d\right\rangle$ where $\alpha^*\alpha$ is the probability for up and $\beta^*\beta$ the probability for down. Because a phase change $\alpha, \beta \to e^{i\vartheta}\alpha, e^{i\vartheta}\beta$ does not change the probabilities and the total probability satisfies $\alpha^*\alpha + \beta^*\beta = 1$, there are only two degrees of freedom. The measured value for the spin of an electron for $\alpha = \beta$ is $\pm\frac{1}{2}$, but the average is 0. (The averages behave classically.) The state vector $(1/\sqrt{2})(\left|u\right\rangle + \left|d\right\rangle)$ points in $x$-direction and $(1/\sqrt{2})(\left|u\right\rangle - \left|d\right\rangle)$ points opposite to the $x$-direction. The state vector $(1/\sqrt{2})(\left|u\right\rangle + i \left|d\right\rangle)$ points in $y$-direction and $(1/\sqrt{2})(\left|u\right\rangle - i \left|d\right\rangle)$ points opposite to the $y$-direction.

Particles and antiparticles have always the same spin. An object made out of an even number of fermions gives a boson (and can have a spin greater than 2). The hydrogen atom, for example, is a boson with spin 1, and the deuterium is a fermion with spin $\frac{3}{2}$. This leads to an kind of paradox. Two fermions make a boson, but then the question arises how one can put them together into the same state.

## 1.12 Symmetric and Antisymmetric Wave Functions

Because the wave function for two bosons $b_1$ and $b_2$ is symmetric such that $\varphi(b_1, b_2) = \varphi(b_2, b_1)$, the real wave function describing the state of the system consisting of these two bosons is the symmetrized wave function $\varphi(b_1, b_2) + \varphi(b_2, b_1)$. Because the wave function for two fermions $f_1$ and $f_2$ is antisymmetric such that $\psi(f_1, f_2) = -\psi(f_2, f_1)$, the real wave function describing the state of the system consisting of these two fermions is the antisymmetrized wave function $\psi(f_1, f_2) - \psi(f_2, f_1)$. Therefore, the two wave functions anticommute. Symmetrizing and antisymmetrizing guarantees the given condition.

Two bosons in the same quantum state can be written as $\varphi(b_1)$ and $\varphi(b_2)$ and the corresponding two-particle state is $\varphi(b_1)\,\varphi(b_2)$. This is obviously symmetric because wave functions are not operators but commute such that $\varphi(b_1)\,\varphi(b_2) = \varphi(b_2)\,\varphi(b_1)$. Thus, two bosons can be in the same state. Two fermions on the other hand cannot be in the same state because $\psi(f_1)\,\psi(f_2) \neq -\psi(f_2)\,\psi(f_1)$.

Four fermions such as two electrons $e_1$ and $e_2$ plus two protons $p_1$ and $p_2$ build two hydrogen atoms $\varphi_1(e_1, p_1)$ and $\varphi_2(e_2, p_2)$ at different locations. The wave function of the individual hydrogen atoms is antisymmetric, but the system of the two hydrogen atoms together is symmetric. The wave function can be written as $\varphi_1(e_1,p_1)\varphi_2(e_2,p_2) - \varphi_1(e_2,p_1)\varphi_2(e_1,p_2) - \varphi_1(e_1,p_2)\varphi_2(e_2,p_1) + \varphi_1(e_2,p_2)\varphi_2(e_1,p_1)$ describing the system of the two hydrogen atoms. The Pauli principle creates a kind of force, and the two atoms cannot be put into the same position state, but they can both be in the same momentum state. This is symmetric with respect to the exchange of the two atoms. Both atoms in the same state means $2\varphi(e_1,p_1)\varphi(e_2,p_2) - 2\varphi(e_2,p_1)\varphi(e_1,p_2) \neq 0$, but this leads to $\varphi(e,p_1)\varphi(e,p_2) - \varphi(e,p_1)\varphi(e,p_2) = 0$ if the two electrons are in the same state.

## 1.13 Half Integer Spin

Spin $S$ and (orbital) angular momentum $L$ are mathematically the same such that $[S_x, S_y] = i\hbar S_z$ and $S_z = m\hbar$, and the operators $S_\pm = S_x \pm iS_y$ can be defined as introduced above. (In the following $\hbar$ is set to 1.) The spin $\frac{1}{2}$ can assume one of the two values $\frac{1}{2}$, $-\frac{1}{2}$, and the spin $\frac{3}{2}$ can be in one of the four states $\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$. Electrons, quarks and neutrinos have spin $\frac{1}{2}$, which is therefore of special interest.

The spin matrices $S_i$ are the same as the Pauli matrices except for a factor $\frac{1}{2}$ such that $S_i = \frac{1}{2}\sigma_i$. They are not unique because their choice depends on the orientation of the axes. The $2 \times 2$ matrices

$$S_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad S_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad S_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.12}$$

are the spin operators selected as usual with the preferred direction $z$ and therefore with

$$|u\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |d\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

such that the two orthonormal states $|u\rangle$ for up and $|d\rangle$ for down build the basis of eigenvectors of $S_z$ with the corresponding eigenvalues $\pm\frac{1}{2}$. The equations

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \pm\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \qquad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \pm\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} -i\beta \\ i\alpha \end{pmatrix}$$

deliver

$$|r\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle) \qquad |l\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|u\rangle - |d\rangle)$$
$$|i\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} = \frac{1}{\sqrt{2}}(|u\rangle + i\,|d\rangle) \qquad |o\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} = \frac{1}{\sqrt{2}}(|u\rangle - i\,|d\rangle)$$

the two eigenvectors $|r\rangle$ and $|l\rangle$ of $S_x$ as well as the two eigenvectors $|i\rangle$ and $|o\rangle$ of $S_y$. The probability for these states is $\frac{1}{2}$ each being $|u\rangle$ or $|d\rangle$.

## 1.14 Integer Spin

The spin 0 can only be in the state 0, the spin 1 can be in one of the three states 1, 0, $-1$, and the spin 2 can be in one of the five states 2, 1, 0, $-1$, $-2$. Only spin 1 will be examined here which corresponds to a three-dimensional spin space. Spin 1 has the three eigenstates $|u\rangle$, $|n\rangle$, $|d\rangle$ corresponding to the three eigenvalues $-1$, 0, 1, respectively. The $3 \times 3$ matrices (check correctness of the commutator relations!)

$$S_x = i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & +1 \\ 0 & -1 & 0 \end{pmatrix} \qquad S_y = i\begin{pmatrix} 0 & 0 & +1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \qquad S_z = i\begin{pmatrix} 0 & +1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

satisfy $[S_x, S_y] = iS_z$ and are possible spin operators, and the eigenvectors of $S_z$, for example, are

$$i\begin{pmatrix} 0 & +1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = m\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = i\begin{pmatrix} \beta \\ -\alpha \\ 0 \end{pmatrix} \quad \Rightarrow \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{or} \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix}$$

for the eigenvalues $-1$, 0, $+1$, respectively. The other eigenvectors can be determined similarly.

## 1.15 Spin and Other Observables

A real electron or any other particle has not only spin, but spin and either position or momentum. The spinners can be written in components

$$|\psi\rangle = \psi(x) = \begin{pmatrix} \psi_u(x) \\ \psi_d(x) \end{pmatrix} \qquad |\varphi\rangle = \varphi(x) = \begin{pmatrix} \varphi_u(x) \\ \varphi_n(x) \\ \varphi_d(x) \end{pmatrix}$$

as shown for a fermion with spin $\frac{1}{2}$ on the left side and a boson with spin 1 on the right side. The probability of an electron (as a fermion) at position $x$ is $\psi_u^*(x)\psi_u(x)$ for up and $\psi_d^*(x)\psi_d(x)$ for down. These probabilities can depend on the position $x$. An electron is completely described by a wave function that has two components and either depends on position or on momentum. Similarly, bosons with spin 1 are completely described by wave functions with three components. The probability to find the particle in the state with spin 1 is $|\langle m=1|\varphi\rangle|^2 = \left|(1/\sqrt{2})\,(1\varphi_u(x) - i\varphi_n(x))\right|^2 = (\varphi_u(x)^2 + \varphi_n(x)^2)/2$ using the above found eigenvector $\langle m=1| = (1/\sqrt{2})\,(1, -i, 0)$. Thus, the total probability to measure either value $-1$, 0 or $+1$ of the spin is $(\varphi_u(x)^2 + \varphi_n(x)^2)/2 + \varphi_d(x)^2 + (\varphi_u(x)^2 + \varphi_n(x)^2)/2 = 1$.

## 1.16 Spin and the Dirac Equation

The Dirac equation and possible matrices $\alpha_j$ and $\beta$ are

$$i\frac{\partial}{\partial t}\Psi = -i\alpha_j\frac{\partial}{\partial x^j}\Psi + \beta m\Psi \qquad \alpha_j = \begin{pmatrix} \sigma_j & 0 \\ 0 & -\sigma_j \end{pmatrix} \qquad \beta = \begin{pmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix}$$

as shown in (1.10) and (1.11), though another set of Dirac matrices is used here. Without $\beta$, the Pauli matrices would solve the problem, but $4\times4$ matrices are needed with $\beta$. The fermions can have positive or negative energy and spin $+\frac{1}{2}$ and $-\frac{1}{2}$. This gives four entries in the matrices.

If the particle is at rest, $k$ as the momentum is zero, and the terms with the matrices $\alpha_j$ disappear. The system can be described by

$$i\frac{\partial}{\partial t}\Psi = \beta m\Psi \qquad \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \end{pmatrix} = \begin{pmatrix} \Psi_+ \\ \\ \Psi_- \end{pmatrix} \qquad \begin{matrix} \Psi_+ = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix} \\ \Psi_- = \begin{pmatrix} \Psi_3 \\ \Psi_4 \end{pmatrix} \end{matrix} \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \Psi_+ \\ \Psi_- \end{pmatrix} = \begin{pmatrix} \Psi_- \\ \Psi_+ \end{pmatrix}$$

showing $i\partial_t\Psi_+ = m\Psi_-$ and $i\partial_t\Psi_- = m\Psi_+$. By adding and subtracting $\Psi_+$ and $\Psi_-$ the Dirac equations become $i\partial_t(\Psi_+ + \Psi_-) = m(\Psi_+ + \Psi_-)$ and $i\partial_t(\Psi_+ - \Psi_-) = -m(\Psi_+ - \Psi_-)$ which are two uncoupled equations. Because energy is just the frequency $\omega$ for $\hbar = 1$, this corresponds to $\omega = \pm m$ and therefore to particles with positive and negative energy. The Dirac equation predicts spin as well as positive and negative energy particles. The split into $\Psi_+$ and $\Psi_-$ corresponds to the two spin states, and the split into $\Psi_+ + \Psi_-$ and $\Psi_+ - \Psi_-$ corresponds to positive and negative energy, respectively.

## 1.17 Lagrangians

Boson fields, fermion fields, the electromagnetic field and a system of fields can all be described by one function called Lagrangian $\mathcal{L} = \mathcal{L}(\Phi, \partial_t\Phi, \partial_x\Phi, \partial_y\Phi, \partial_z\Phi) = \mathcal{L}(\Phi, \partial_\mu\Phi)$ with the equation of motion

$$\frac{\partial}{\partial t}\frac{\partial\mathcal{L}}{\frac{\partial}{\partial t}\Phi} + \frac{\partial}{\partial x}\frac{\partial\mathcal{L}}{\frac{\partial}{\partial x}\Phi} + \frac{\partial}{\partial y}\frac{\partial\mathcal{L}}{\frac{\partial}{\partial y}\Phi} + \frac{\partial}{\partial z}\frac{\partial\mathcal{L}}{\frac{\partial}{\partial z}\Phi} = \frac{\partial\mathcal{L}}{\partial\Phi} \tag{1.13}$$

for the field $\Phi$. The simplest example is a scalar field with the Lagrangian

$$\mathcal{L} = \frac{1}{2}\left(\frac{\partial \Phi}{\partial t}\right)^2 - \frac{1}{2}\left(\frac{\partial \Phi}{\partial x}\right)^2 - \frac{1}{2}\left(\frac{\partial \Phi}{\partial y}\right)^2 - \frac{1}{2}\left(\frac{\partial \Phi}{\partial z}\right)^2 - V(\Phi) \qquad\qquad V(\Phi) = \frac{m^2\Phi^2}{2}$$

where $V(\Phi)$ is the potential where one example for such a potential is shown. Constant potentials do not contribute, and linear potentials lead to unbound negative energies, thus the simplest possible potential is proportional to $\Phi^2$. The equation of motion is

$$\frac{\partial^2 \Phi}{\partial t^2} - \frac{\partial^2 \Phi}{\partial x^2} - \frac{\partial^2 \Phi}{\partial y^2} - \frac{\partial^2 \Phi}{\partial z^2} = -\frac{\partial V}{\partial \Phi} \qquad\qquad \frac{\partial^2 \Phi}{\partial t^2} - \frac{\partial^2 \Phi}{\partial x^2} - \frac{\partial^2 \Phi}{\partial y^2} - \frac{\partial^2 \Phi}{\partial z^2} = -m^2\Phi$$

in general and for the example potential. That is a very basic form of a wave equation both in classical field theory, Quantum Field Theory, electrodynamics and so on.

In Quantum Mechanics with the field $\Phi = e^{-i\omega t} \cdot e^{+ikx}$ the usual relation between energy, mass and momentum is $E^2 - p^2 = m^2$ or $\omega^2 - k^2 = m^2$ with $c = \hbar = 1$ coming from $-\omega^2\Phi + (k_x^2 + k_y^2 + k_z^2)\Phi = -m^2\Phi$ where the value $m$ is interpreted as the mass of a quantum. This is invariant under Lorentz transformation. (A general rule is that the only thing one has to do to ensure invariance under Lorentz transformations is to make sure that the Lagrangian is a scalar.)

If the differential equations are linear equations, two different solutions can be added to give another solution. If one adds, for example, the term $g\Phi^3$ to the potential, the non-linear term $-3g\Phi^2$ appears in the equation of motion, and non-linearity creates scattering of waves.

The Lagrangian of two scalar fields $\Phi_1 + \Phi_2$

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi_1)^2 + \frac{1}{2}(\partial_\mu \Phi_2)^2 - \frac{m_1^2\Phi_1^2}{2} - \frac{m_2^2\Phi_2^2}{2}$$

does not encode any interaction between the two fields because the field $\Phi_2$ does not appear in the equation of motion for $\Phi_1$ and vice versa. An interaction term must have the form such as $-\Phi_2\Phi_1^2$ leading to a term $-2\Phi_2\Phi_1$ in the equation of motion for $\Phi_1$ and to a term $-\Phi_1^2$ in the equation of motion for $\Phi_2$ to produce interactions.

Also the Dirac field can be encoded in a Lagragian which has basically the form

$$\Psi^\dagger \frac{\partial}{\partial t}\Psi + \Psi^\dagger \alpha \frac{\partial}{\partial t}\Psi + \Psi^\dagger \beta \Psi m$$

with the Dirac matrices (1.10). A term of the form $\Psi^\dagger\beta\Psi\Phi$ couples a Dirac field $\Psi$ with a scalar field $\Phi$ such that the Dirac field scatters the scalar field and the scalar field scatters the Dirac field. The Lagrangian for the coupled fields produces the equation of motion for both fields.

Quantum fields are built up of creation and annihilation operators. Cubic terms and terms with higher powers therefore contain several creation and annihilation operators and mean some particles coming in and some particles going out at a given point in space. The term $\Psi^3(x)$, for example, creates and/or annihilates three particles at point $x$. It may absorb one particle and emit two, it may absorb all three and so on. These processes can be described by Feynman diagrams.

Quadratic terms on the other hand govern the motion of undisturbed particles. A term of the form

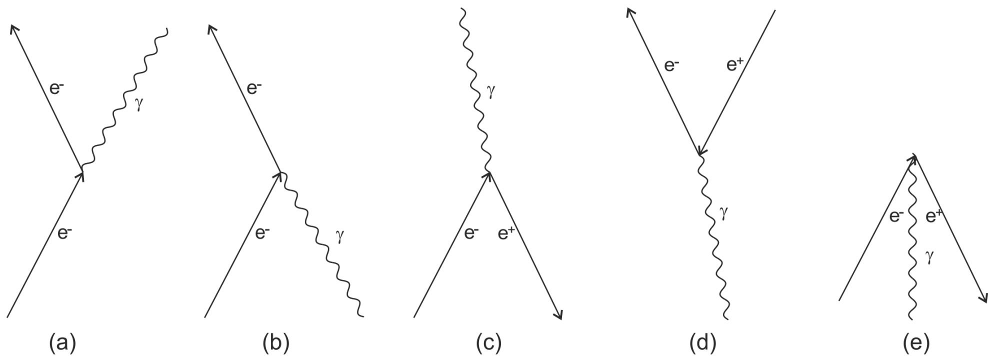
$$\frac{1}{2}\left(\frac{\partial}{\partial x}\Phi(x)\right)^2 = \frac{(\Phi(x) - \Phi(x'))^2}{2\varepsilon^2} = \frac{1}{2}\frac{\Phi(x)^2}{\varepsilon^2} + \frac{1}{2}\frac{\Phi(x')^2}{\varepsilon^2} - \frac{\Phi(x)\Phi(x')}{\varepsilon^2}$$

shows the field at two neighboring points $x$ and $x'$ with the infinitesimal distance $\varepsilon$. The terms $\Phi(x)^2$ and $\Phi(x')^2$ as well as the massterm $m^2\Psi^2(x)$ encode the annihilation and creation of a particle at the same point in space and do not contribute to the motion. The cross-term $\Phi(x)\Phi(x')$, however, absorbs the particle at $x$ and reemits it at $x'$ and therefore moves the particle from $x$ to $x'$.

The Lagrangians encode these processes in a condensed form, and they are known in Particle Physics to a high precision through experiments. They play on one side a much more important role in Quantum Physics than in classical physics, and taking on the other side a Lagrangian from Quantum Mechanics for a large number of particles gives the classical Lagrangian.

## 1.18 Feynman Diagrams

Feynman diagrams allow to visualize processes in which fields interact with each other. In the figure below, processes from Quantum Electrodynamics are illustrated. The five diagrams show (a) an electron $e^-$ emitting a photon $\gamma$, (b) an electron $e^-$ absorbing a photon $\gamma$, (c) the annihilation of a pair of an electron $e^-$ and a positron $e^+$ into a photon $\gamma$, and (d) the decay of a photon $\gamma$ into an electron $e^-$ and a positron $e^+$. Also processes which are not allowed can be drawn such as (e) the annihilation of an electron $e^-$, a positron $e^+$ and a photon $\gamma$ violation energy conservation.



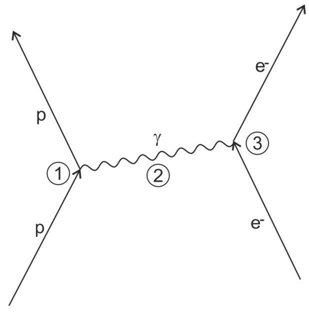

In the operators $\Psi^\dagger$ and $\Psi$ of the Dirac equation are creation and annihilation operators for positive and negative energy. Because removing a particle with negative energy is adding an antiparticle with positive energy, terms $c^\pm$ for negative energy can be replaced by $c^\mp$ for positive energy

$$\Psi = c_{\mathrm{e}}^- + c_{\mathrm{p}}^+ \qquad \Psi^\dagger = c_{\mathrm{e}}^+ + c_{\mathrm{p}}^- \qquad \Psi^\dagger(x)\Psi(x)A_0 + \Psi^\dagger(x)\alpha\Psi(x)A_i$$

and $\Psi$ increases the electric charge by one unit either through the annihilation of an electron with $c_{\mathrm{e}}^-$ or through the creation of a positron with $c_{\mathrm{p}}^+$ and $\Psi^\dagger$ decreases it by one unit either through the creation of an electron with $c_{\mathrm{e}}^+$ or the annihilation of a positron with $c_{\mathrm{p}}^-$. The part of the Lagrangian containing the interaction terms include the photon field $A$ (creation and annihilation operators for photons) with the electromagnetic vector potential $A_\mu$. All processes shown above as Feynman diagrams are described by the above term in the Lagrangian. The factor $e$ in front of this term connects the probability to the electric charge which is a dimensionless quantity measured in multiples of the charge of an electron. The Feynman diagrams show conservation of charge. Conservation of energy needs integration over all times, and conservation of momentum needs integration over the whole space. Doing so would show that the process (e) in the above figure is not possible.

Processes formed by systems of protons, neutrons and mesons are fairly similar to processes in Quantum Electrodynamics, and the corresponding Feynman diagrams look therefore also very similar to the ones in the above figure. Protons, neutrons and mesons are, however, not elementary particles because they are built up from quarks. If one hits protons and neutrons hard enough, they would split up into quarks although free quarks have never been observed. Protons and neutrons consist of three quarks, and mesons consist of a quark and an antiquark. One interesting kind of mesons are pions $\Pi_+$, $\Pi_-$ and $\Pi_0$ where one has positive charge, one has negative charge and one is electrically neutral. Protons and neutrons are fermions represented by Dirac fields $\Psi_{\mathrm{p}}$ and $\Psi_{\mathrm{n}}$, respectively. The proton is positively charged, and the neutron is electrically neutral.

The process $\Psi_{\mathrm{p}}^\dagger\Psi_{\mathrm{n}}\Pi_+$ shows a neutron and a positively charged pion coming in and a positron going out. Similarly, $\Psi_{\mathrm{n}}^\dagger\Psi_{\mathrm{p}}\Pi_-$ shows the process where a proton and a negatively charged pion are absorbed leading to the emission of a neutron. Another example is the absorption of an proton and an antineutron emitting a positively charged pion. This was the view when protons, neutrons and mesons are believed to be elementary particles.




The process where a proton emits a photon which is later absorbed by an electron results in a product of many Lagrangians at different points in space because single Lagrangians describe processes at one point or at two very near points. The emission of the photon at ① and the absorption of the photon at ③ are each described by one Lagrangian. The motion of the photon at ② needs very many Lagrangians to move the photon from point to neighboring point because it has to be annihilated at a point and created at a close point.

Fields have operators and are symbols which make processes happen. Elementary processes emit a particle or absorb a particle, and therefore a single field by itself can either absorb a particle or emit a particle. The product of two fields can absorb a particle and emit a particle. In terms of spacetime one thinks of a trajectory as being built up in many little steps leading to the motion of the particle. To model the transition from a point to a very close neighboring point, one breaks space up into a lot of little cells. Quantum Field Theory is the limit of a theory with finite cells to continuous spacetime.

## 1.19 More on Lagrangians

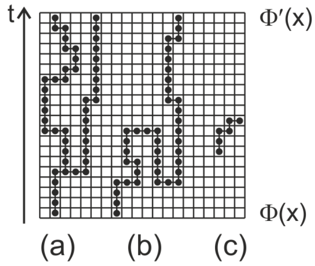
A general process has the form $\langle 0|\Psi^\dagger ... \Psi^\dagger\, \mathcal{L} ... \mathcal{L}\, \Psi ... \Psi|0\rangle$ where $\Psi ... \Psi\,|0\rangle$ is the initial state and $\langle 0|\, \Psi^\dagger ... \Psi^\dagger$ the final state. For electrons, an equal number of $\Psi$ and $\Psi^\dagger$ means conservation of charge. The art is to find the Lagrangian from experimental data. Once it is found, the rules are very definite and lead in a precise way to the probabilities. In the physicists mind, there is one Lagrangian that governs all of nature. Every time a new particle is discovered a new field has to go into this Lagrangian, and every time a new process is found a new interaction has to be added to this Lagrangian. If one would write down the Lagrangian of all known physics, this would cover a couple of closely spaced pages. To study Quantum Electrodynamics, for example, one does not need the whole big Lagrangian if one does not need infinite precision, and one can, for example, ignore quarks because they do not heavily interact with electrons. Thus, one can study a small part of the big Lagrangian. Lagrangians must, however, be relativistic and be in accordance with Special Relativity and Quantum Mechanics.

The path integral formalism is a generalization and the quantum mechanical version of the principle of least action. This principle exists not only in classical mechanics where the action is the Lagrangian integrated over time but also in classical field theory where the action is the Lagrangian (which depends on the field $\Phi$ and the derivatives $\partial_\mu \Phi$) integrated over spacetime between the initial and final configuration. Special Relativity requires the Lagrangian $\mathcal{L}$ to be a scalar. The trajectory becomes the amplitude to find a particle which was initially at a given position at a final position in Quantum Mechanics. Feynman's rule is that the sum (or integral) $\sum e^{-i/\hbar \int \mathcal{L} dt}$ over all possible trajectories from $(x, t)$ to $(x', t')$ is the amplitude to go from one point to another. To use this in Quantum Field Theory, the same idea as in classical field theory is used. One starts from an initial configuration of the field and determines the probability amplitude to find it at a later time in a given final configuration. This is the sum (or integral) $\sum e^{-i/\hbar \int \mathcal{L}\, dt\, dx\, dy\, dz}$ summed over all possible histories. Because of the duality between particles and fields, the initial configuration can be described using the incoming particles as the field quanta instead of the values of the field and similarly for the final configuration using the outgoing particles.

## 1.20 Lattice Gauge Theory

To determine the result for the Feynman path integration, the continuous spacetime is discretized into a lattice of cells of finite size. At the end, the limit for the size of the cells getting smaller and smaller is taken to go back to the continuous spacetime. The terms $e^{-i/\hbar \int \mathcal{L}\, dt\, dx\, dy\, dz}$ can be replaced by $e^{-i/\hbar \sum A_j}$ where the $A_j$ are the actions in the cells, and this can be replaced by $\prod e^{-i/\hbar A_j}$ or with $\prod e^{-iA_j}$ when setting $\hbar = 1$. For a quantum field the initial configuration is the values of the field in the first row $\Phi(x)$ and the final configuration is the field values in the last row $\Phi'(x)$. To get the amplitude for the initial configuration to become the final configuration, all the $e^{-iA_j}$ in the cells of a history may be replaced by $1 - iA_j$ because $e^\varepsilon \approx 1 + \varepsilon$ for small values of $\varepsilon$ and are multiplied together. Fields are creation and annihilation operators, and the product of the fields in a cell times the fields in a neighboring cell can represent the annihilation of a particle in the first cell and the creation of a particle in the other cell. This could be represented by a particle moving from one cell the other.

Thus, derivatives involve two cells in the calculations above such that also terms of the form (with some coefficients left out) $e^{-i \sum \Phi(x)\Phi(x')}$ are needed in the action where $x$ and $x'$ are neighboring points (neighboring cells in the lattice). Derivatives with respect to time are vertical (from initial to final configuration), and derivatives with respect to space are horizontal (parallel to the initial and final configuration). Using the Taylor expansion $e^{-i\sum \Phi(x)\Phi(x')} \approx 1 - i\sum \Phi(x)\Phi(x') - \frac{1}{2!}(\sum \Phi(x)\Phi(x') \sum \Phi(x'')\Phi(x'''))$ and so on, the first term 1 can be ignored, and the second term shows the movement of a particle from the cell $x$ to the neighboring cell $x'$. The third term encodes two movements one from $x$ to $x'$ and one from $x''$ to $x'''$. These movements only contribute to the amplitude if they do not leave dangling ends except for the initial and final particles. Thus, the second term in the Taylor expansion only contributes if the movement from a cell in $\Phi(x)$ leads to a cell in $\Phi'(x)$, and similarly, the third term alone only avoids dangling ends if $x$ belongs to $\Phi(x)$, $x' = x''$ and $x'''$ belongs to $\Phi'(x)$ and if particles can go from $\Phi(x)$ to $\Phi'(x)$ in two steps.



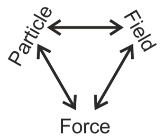

Higher-order terms of the Taylor expansion allow longer paths from a cell belonging to $\Phi(x)$ to a cell belonging to $\Phi'(x)$. All the possible routes from a specific cell $x_\mathrm{i}$ in $\Phi(x)$ to a specific cell $x_\mathrm{f}$ in $\Phi'(x)$ contribute to the amplitude for going from $x_\mathrm{i}$ to $x_\mathrm{f}$. The figure on the right side shows three cases. Situation (a) depicts a particle moving up due to quadratic terms in the Lagrangian which annihilate a particle in one cell and create a particle in a neighboring cell until a cubic term $g\Phi^3$ in the Lagrangian annihilates one particle and creates two particles which appear in the final configuration. Situation (b) shows an original particle moving up and a pair of particle and antiparticle being created. The original particle and the antiparticle annihilate, and the newly created particle continues. Dangling endpoints as in (c) are not allowed except if there is explicit instructions to create a particle in the lower cell and to annihilate it in the upper cell.

The Feynman rules and the Lagrangian contain the same information. The Feynman path integral corresponds to the sum over all possible paths from the initial to the final configuration, and the coupling constants in the Lagrangian make sure that the sum converges. The higher terms get higher powers of the coupling constants which are usually small numbers. Quantum Field Theory has beautiful aspects, but writing down all the possible processes is just a mess. There are many masses, coupling constants and so on. But when these numbers are measured, one can describe things to a great precision. Quantum Electrodynamics is the simplest theory because all there is are electrons and photons, and the coupling constant is the electric charge $e$. Scattering a photon by an electron and scattering an electron by a photon are very improbable events because the probability $e^4$ is a very small number. For every more complicated situation an additional factor $e^2$ comes in. This makes sure that the simpler Feynman diagrams contribute more than the complicated ones.

Lattice gauge theory or lattice quantum field theory has taken some thirty years by now and has become a big industry. A lot of the wisdom came from statistical mechanics because similar problems occur there. The lattice could represent a real crystal, for example, instead of a field.

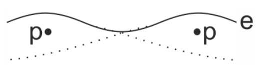
# 2 Standard Model

## 2.1 Particles, Fields and Forces




Fields and waves are more or less the same things, and elementary particles are their quanta. Forces can be seen from the field's point of view, but they can also be seen from the particle's point of view, because a field is a collection of quanta. This can be shown for the electromagnetic field.

Two particles with charge $e_1$ and $e_2$, respectively, build together the total electric field $\vec{E} = e_1\vec{E}_1 + e_2\vec{E}_2$ with the energy $E_N = \int e_1^2\vec{E}_1^2 + e_2^2\vec{E}_2^2 + 2\,e_1e_2\,\vec{E}_1 \cdot \vec{E}_2\,dV$ where the first and second term are the two fields alone and the third term corresponds to the Coulomb force. Bringing the two particles far away makes the third term very small. If one brings them together, this distorts the field between them. This is a pure field view of forces, and it is based on classical physics.




Two protons are assumed to be far but not infinitely far away from each other. The ground state is the same if an electron orbits around one or the other proton. If the two protons are not infinitely far away, there are processes possible which would not be possible if they were infinitely far away. Quantum Mechanics allows the electron to tunnel and therefore to hop back and forth with a small probability. It has to overcome an energy barrier in between which would not be possible to overcome in classical physics but which is possible in Quantum Mechanics because of the tunneling effect. Long-term an equilibrium is reached where the electron orbits around either proton with equal probability. The energy of the electron is not exactly the same when the two protons are close together as it is when the two protons are infinitely far away. Bringing the two protons closer together changes the energy. Plotting the energy as a function of the distance between the protons shows a certain energy level for large distances but a slightly lower energy level, when the protons are close together. Because forces move objects always toward lower energy, there is an attraction force between two protons with an electron hoping back and forth, and this force is called particle exchange force. It can therefore happen that electrons tunnel between two hydrogen atoms.

There is a concept which shows up in classical electrodynamics that the presence of two charges changes the forces compared to the presence of only one, but which also shows up in Quantum Mechanics between two hydrogen atoms. Being in a superposition of states lowers the energy and creates a force. In Quantum Electrodynamics a proton is interacting with the electromagnetic field whose quanta are photons. The force between two protons results from the emission and absorption of photons. In Quantum Field Theory, all forces are exchanges of particles. The electron, for example, is a superposition of a state with no photon, a state where a photon has been emitted, a state where two photons have been emitted, a state where a photon has been reabsorbed and so on. If there is another electron close by one electron can absorb a photon emitted by the other electron. If one determines the energy of the two electrons with respect to their distance, the two energies just add for very large distance between the two electrons. If they come closer and closer, the charges begin to influence each other and the exchange of photons creates a force. Any particle can be exchanged in some context or another. Thus, any particle produces a force. In molecular physics the exchange of electrons creates forces.

There is a one-to-one correspondence between particles and fields for which the particles are the quanta, and a one-to-one correspondence between particles and forces which are associated with exchange processes where that particular particle can jump back and forth. Therefore there are not four forces in nature but one force for every kind of particle.

## 2.2 Some Elementary Particles

The elementary particles are grouped together for better understanding, but still there are more questions than answers. Nobody knows why particles such as the electron exist but others do not, or why certain quantities have the values they do. The masses of the particles, for example, are a puzzle, and nobody understands why they are the way they are. The numbers can be measured although different sources show slightly different values, but there is no theory which would explain exactly the values measured. Also the names for the groups were invented in the past, and some names of the groups such as the name mesons are today misleading because of new insight. Some relationships are known, but there are many more parameters and many more different types of particles than there are known relationships between them. Particle physics is a big mess.

As the first particle, the photon is mentioned here. It has no mass and no electric charge. It is a boson, and it is its own antiparticle. The symbol of the photon is $\gamma$, and the corresponding field is the vector potential of the electromagnetic field denoted by $A$. Its spin is 1, and the baryon number is 0.

The electron with the symbol $e^-$ and the electric charge $-1$ and its antiparticle positron with the symbol $e^+$ and the electric charge $+1$ are fermions, and their field is denoted by $\Psi_e$. The mass of the electron is $0.511\,\mathrm{MeV}$, it has spin $\frac{1}{2}$, and the baryon number is 0.

Quantum Electrodynamics describes the interaction between photons, electrons and positrons. The basic process is the emission of a photon by an electron, and the term $e\Psi_e^\dagger\Psi_e A$ in the Lagrangian represents all the possibilities where an electron or a positron emits or absorbs a photon and where a photon decays into an electron and a positron or where an electron and a positron annihilate with the emission of a photon. These interactions build the simplest group of Feynman diagrams. If the nucleus is seen as a point, Quantum Electrodynamics is also the theory of the atoms.

The proton $p$ or $p^+$ and the neutron $n$ building the nucleus of an atom are not elementary, and their baryon number is 1. The antiproton $p^-$ or $\overline{p}$ and the antineutron $\overline{n}$ have therefore baryon number $-1$. They all consist of quarks. The symbol $q$ and $\Psi_q$ denotes quarks in general and their fields. They are fermions, have baryon number $\frac{1}{3}$ and spin $\frac{1}{2}$. There are six different quarks, each with its own antiquark which have the same electric charge as the given quark except for the opposite sign as it is normal for antiparticles. Their name, symbol, antiquark, charge, and mass are:

| | | | | |
|---|---|---|---|---|
| up | $u$ | $\overline{u}$ | 2/3 | 2.2 MeV |
| down | $d$ | $\overline{d}$ | −1/3 | 4.7 MeV |
| charm | $c$ | $\overline{\mathrm{c}}$ | 2/3 | 1 280 MeV |
| strange | $s$ | $\overline{\mathrm{s}}$ | −1/3 | 96 MeV |
| top | $t$ | $\overline{\mathrm{t}}$ | 2/3 | 173 100 MeV |
| bottom | $b$ | $\overline{\mathrm{b}}$ | −1/3 | 4 180 MeV |

Quarks which are themselves not observable particles build the nuclei. The proton is $uud$ and has therefore electric charge 1, and the neutron is $ddu$ and has therefore electric charge 0. Protons and neutrons are very similar with the exception of the electric charge. The neutron with a mass of 939.57 MeV is a little bit heavier than the proton with a mass of 938.27 MeV. The mass of the two nuclei comes from the masses of the three quarks with the gluons carrying no mass but energy and keeping them together. Down quarks can be replaced by strange or a bottom quarks, up quarks can be replaced by charm or top quarks, and the nuclei become heavier but unstable. The number of quarks minus the number of antiquarks is conserved, but the question is whether this is an absolute conservation law. Without violating charge conservation a proton could decay into a positron and a photon and thus reduce the baryon number violating baryon number conservation. If this process happens then it has a very small probability, and whether it happens is one of the very interesting questions of Particle Physics.

Also mesons are not elementary particles because they consist of a quark and an antiquark. They have baryon number 0. The two pions $\Pi^+$ and $\Pi^-$ being antiparticles of each other are $u\overline{d}$ and $\overline{u}d$, and $\Pi^0$ is the superposition $(1/\sqrt{2})(|u\overline{u}\rangle - |d\overline{d}\rangle)$. Except for their electrical charge and a tiny difference in mass, they are nearly indistinguishable, and their mass is approximately 140 MeV. A state $(1/\sqrt{2})(|u\overline{u}\rangle \pm |d\overline{d}\rangle)$ is an entangled pair of quarks not entangled through spin but through the so-called isospin. Other mesons such as the $\eta$ corresponding to $(1/\sqrt{2})(|u\overline{u}\rangle + |d\overline{d}\rangle)$ with a mass in the order of 500 MeV are heavier than the pions which are also called $\pi$-mesons. There are many other mesons including four kaons which are also called K-mesons and which contain a strange quark or antiquark. The electrically neutral $\mathrm{K}^0$ with the antiparticle $\overline{\mathrm{K}}^0$ is $d\overline{s}$, and the electrically charged $\mathrm{K}^+$ with the antiparticle $\mathrm{K}^-$ is $u\overline{s}$. Mesons are all bosons, and they are unstable with average lifetime in the order of $10^{-8}$ s or less.

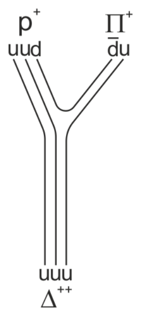
## 2.3 Quantum Chromodynamics

Quantum Chromodynamics is the theory of quarks and gluons. Similar to the electromagnetic force which is seen as an exchange of virtual photons in Quantum Electrodynamics, the electrically neutral gluons act as exchange particles for the strong force between quarks in Quantum Chromodynamics.

Mathematically, spin and isospin (also called isotopic spin) are isomorphic. To distinguish isospin written as $(u, d)$ from spin, the spin is denoted by $(\uparrow, \downarrow)$. The isospin interchanges up quarks and down quarks (protons and neutrons). This is not a precise symmetry because one has to ignore the difference of mass of the up and down quarks (proton and neutron), and one has to ignore also electric charge.

For isospin $\frac{1}{2}$, there are the two states proton and neutron. Because quarks are fermions, the states have to be antisymmetrized such that $p^+ = d_1(u_2u_3 - u_3u_2)$ for the proton and $n = u_1(d_2d_3 - d_3d_2)$ for the neutron. The antisymmetrized parts in the bracket have isospin 0, and the total isospin is $\frac{1}{2}$.

For isospin $\frac{3}{2}$, there are the two particles $\Delta^{++}$ with $uuu$ and $\uparrow\uparrow\uparrow$ or $\downarrow\downarrow\downarrow$ plus charge 2 as well as the $\Delta^-$ with $ddd$ and also $\uparrow\uparrow\uparrow$ or $\downarrow\downarrow\downarrow$ plus charge −1. There are two more states $uud$ and $udd$, which are not proton and neutron. The mass of these particles is around 1200 MeV, and they can decay into protons, neutrons plus some other particles. The $\Delta^{++}$, for example, decays into a proton $p^+$ and a pion $\Pi^+$ because it is not stable.

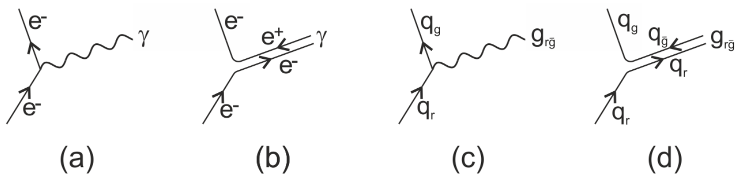



There is something wrong because $uuu$ with $\uparrow\uparrow\uparrow$ is not allowed for fermions. There must be another property which is called color and which gave Quantum Chromodynamics the name. The three possible colors are red, green and blue, but other names are used as well because these colors have nothing to do with ordinary colors but are just labels.

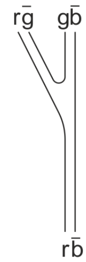
In earlier days because of the Pauli exclusion principle and the helium atom with two electrons in the same state, an at that time unknown property had to exist which turned out to be the spin. The new property color needed for the three $u$ in the same state is a kind of charge which is not measurable. Different quarks must differ in some way, and its not position. Thus, all six flavors of quarks come with all three colors. To distinguish them the color of the up quark, for example, is specified as $u_\mathrm{r}$, $u_\mathrm{g}$ and $u_\mathrm{b}$.

Spin $\uparrow$ and $\downarrow$ are not distinguishable as such (if no direction is given by other means). Only when two particles are there, one can identify whether they have the same spin or opposite spin. The same is true for color. The $\Delta^{++}$ (also called $\Delta_{3/2}$) particle can be written as $u_\mathrm{r}u_\mathrm{g}u_\mathrm{b}$. States such as $|u_\mathrm{r}u_\mathrm{g}u_\mathrm{b}\rangle$ or $|u_\mathrm{r}\overline{u}_\mathrm{g}\rangle$ have to be symmetrized or antisymmetrized depending on whether they are a boson or a fermion.

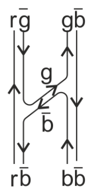
Gluons are bosons which glue quarks together to build nuclei similarly to atoms which are held together by photons. Gluons are very similar to photons. They are massless and have spin 1 with similar polarization states, but there is one big difference. There is no interaction between photons. From a physicist's point of view, they are rather boring. The emission of a photon $\gamma$ by an electron $e^-$ as shown in (a) of the figure below could also be drawn as in (b) where the photon is interpreted as an electron $e^-$ and a positron $e^+$. This picture is not too useful in Quantum Electrodynamics. In Quantum Chromodynamics, the emission of a gluon $g$ could be visualized as in (c), but here the diagram (d) makes more sense.




Gluons always have a color $(\mathrm{r}, \mathrm{g}, \mathrm{b})$ and an anticolor $(\overline{\mathrm{r}}, \overline{\mathrm{g}}, \overline{\mathrm{b}})$ leading to nine combinations. For a gluon with r and $\overline{\mathrm{b}}$, for example, one simply writes $\mathrm{r}\overline{\mathrm{b}}$ instead of $g_{\mathrm{r}\overline{\mathrm{b}}}$. Photons do not interact except in materials, but gluons interact with each other and are therefore much more interesting. Two gluon waves heavily interact when passing each other.




Gluons can emit gluons. This creates a force between two gluons through the exchange of a gluon between them. A single gluon wave influences other parts of the wave. This is not linear as for photon waves. When a gluon of type $\mathrm{r}\overline{\mathrm{b}}$ exchanges a gluon of type $\mathrm{g}\overline{\mathrm{b}}$ with a gluon of type $\mathrm{b}\overline{\mathrm{b}}$ the two original gluons become $\mathrm{r}\overline{\mathrm{g}}$ and $\mathrm{g}\overline{\mathrm{b}}$. This leads to a force between the two original gluons which is similar to the force between a nucleus and an electron based on the exchange of photons. The same kind of mechanism based on the exchange of gluons between quarks is responsible for the strong force within the nuclei.




The mass of objects depends on the frequency. This is also true for quarks and gluons. When three quarks in a proton slowly shake, the whole proton shakes. When they shake with high frequency, only the quarks shake. The mass of quarks is the high-frequency mass.

## 2.4 Rotations and General Symmetry Groups

The colors of the quarks and gluons build a symmetry group corresponding to rotations in an abstract space. Rotations act on vectors, and rotations can be combined such that they have a group structure. Rotations have an axis which is a unit vector, and they have an angle. To describe a rotation, three parameters are therefore needed. The group has an identity, an inverse and a product. Rotations happen to be associative, but they do not commute. There are two kinds of groups, discrete and continuous. Consequently, there are discrete and continuous symmetries. Continuous groups are called Lie groups.

For every axis defined by a unit vector $\hat{n}$ and for every angle $\vartheta$, there is a matrix $\mathbf{R}(\vartheta, \hat{n})$ such that $\sum R_{ij}(\vartheta, \hat{n})v_j = v'_i$ where $\vec{v}$ and $\vec{v}'$ are vectors with components $v_i$ and $v'_i$, respectively. Rotations do not change the length such that $\sum v'_i v'_i = \sum v_j v_j$. From $\sum R_{ij}v_j = v'_i$ and $\sum R_{ik}v_k = v'_i$ follows $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ and therefore $\mathbf{R}^{-1} = \mathbf{R}^T$ because $\sum v'_i v'_i = \sum R_{ij}v_j \sum R_{ik}v_k = v_i v_i$ and therefore $\sum R_{ij}R_{ik} = \delta_{jk}$.

Spin describes angular momentum. One can rotate a spin. Spin 1 particles are called vector particles because spin 1 has three states which can be written as a three-dimensional vector. Spin 0 particles, on the other hand, can be rotated, and nothing happens. For spin $\frac{1}{2}$ particles, the rotation group can be represented by

$$|u\rangle = \begin{pmatrix}1\\0\end{pmatrix} \qquad |d\rangle = \begin{pmatrix}0\\1\end{pmatrix} \qquad \begin{pmatrix}\alpha_1\\\alpha_2\end{pmatrix} = \alpha_1 |u\rangle + \alpha_2 |d\rangle \qquad \begin{pmatrix}u_{11} & u_{12}\\u_{21} & u_{22}\end{pmatrix}\begin{pmatrix}\alpha_1\\\alpha_2\end{pmatrix} = \begin{pmatrix}\alpha_1'\\\alpha_2'\end{pmatrix}$$

where $\alpha_1^*\alpha_1 + \alpha_2^*\alpha_2 = 1 = \alpha_1'^*\alpha_1' + \alpha_2'^*\alpha_2'$. The rotations for spin $\frac{1}{2}$ use complex number consisting of two real numbers each. The fact that the rotation matrices are unitary leaves four independent real numbers, but for a rotation only three degrees of freedom are needed. The fact that the determinant has to be equal 1 reduces the number to three. For the determinant, $\det(\mathbf{U}_1\mathbf{U}_2) = \det(\mathbf{U}_1)\det(\mathbf{U}_2)$ shows that also the product of two rotations with determinant 1 has determinant 1, and because $\mathbf{I}$ has determinant 1 also $\mathbf{U}^{-1}$ has determinant 1. In general, the determinant of a unitary matrix is a phase $e^{i\vartheta}$.

The unitary $2 \times 2$ matrices with determinant 1 build a group with three parameters such as the three Euler angles, for example. It is a rotation group but it acts on the two-dimensional superpositions of $|u\rangle$ and $|d\rangle$. Its name is SU(2) for special unitary group of dimension 2, and it is isomorphic to rotations in three dimensions. Unitary n×n matrices are called U(n), and those with determinant 1 are called SU(n).

With the Pauli matrices $\sigma_j$, the spin operators $S_j$ are $S_j = \frac{1}{2}\sigma_j$. The Pauli matrices themselves are not elements of SU(2), but the matrices $i\sigma_j$ are. Small rotations can be written in the form $\mathbf{U} = \mathbf{I} + \varepsilon\mathbf{M}$. This leads to an antihermitian matrix $\mathbf{M}$. To make it hermitian, one can multiply it by $i$. Thus, because of $\mathbf{U}^\dagger\mathbf{U} = (\mathbf{I} + i\varepsilon\mathbf{M})(\mathbf{I} - i\varepsilon\mathbf{M}^\dagger) = \mathbf{I}$, the matrix is now hermitian as it fulfills $\mathbf{M} - \mathbf{M}^\dagger = 0$, and one can write $\mathbf{I} + i\varepsilon(\mathbf{M} - \mathbf{M}^\dagger) + O(\varepsilon^2) = \mathbf{I}$. The fact that the $m_{jk}$ are very small such that $m_{11}m_{22}$ and $m_{12}m_{21}$ can be ignored in

$$i\varepsilon\mathbf{M} = \begin{pmatrix}1+m_{11} & m_{12}\\ m_{21} & 1+m_{22}\end{pmatrix} \quad\Rightarrow\quad \begin{aligned}\det(i\varepsilon\mathbf{M}) &= (1+m_{11})(1+m_{22}) - m_{12}m_{21}\\ &= 1 + m_{11} + m_{22} + m_{11}m_{22} - m_{12}m_{21} \approx 1 + m_{11} + m_{22}\end{aligned}$$

shows that $\mathbf{M}$ must be traceless. The $\sigma_j$ are the traceless $2 \times 2$ matrices, and all small rotations are therefore a linear combination of the $\sigma_j$. In general, the generators of a group are the infinitesimal elements of the group. One can build any group element out of infinitesimal elements. The generators are conserved quantities, and the conserved quantities associated with a symmetry are the generators.

## 2.5 Symmetry Group of the Colors

Before the detection of quarks involving charm, top, or bottom, the SU(3) was used to describe the approximate symmetry between the quarks with up, down, and strange and to act on them in a similar way as the isospin acts on protons and neurons. After the detection of color, the color SU(3) symmetry has become the standard way of understanding the strong interactions. The color SU(3) is a gauge symmetry and not just an approximation.

The color of quarks is described by $\alpha_1 |r\rangle + \alpha_2 |g\rangle + \alpha_3 |b\rangle$. It adds another symmetry to physics which is described by the group SU(3) containing the unitary $3 \times 3$ matrices with determinant 1. From the nine complex numbers the fact that the matrix is unitary gives nine equations $(\mathbf{U}^\dagger\mathbf{U})_{jk} = \delta_{jk}$ and $\det(\mathbf{U}) = 1$ gives another one. Thus, there are eight degrees of freedom, and that corresponds to the eight gluons. The group SU(3) defines the basic symmetries of Quantum Chromodynamics and its invariances. There must be also eight $3 \times 3$ linearly independent hermitian and traceless matrices. They have been found as

$$\begin{aligned}
&\lambda_1 = \begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix} \quad \lambda_2 = \begin{pmatrix}0&-i&0\\i&0&0\\0&0&0\end{pmatrix} \quad \lambda_3 = \begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix}\\
&\lambda_4 = \begin{pmatrix}0&0&1\\0&0&0\\1&0&0\end{pmatrix} \quad \lambda_5 = \begin{pmatrix}0&0&-i\\0&0&0\\i&0&0\end{pmatrix}\\
&\lambda_6 = \begin{pmatrix}0&0&0\\0&0&1\\0&1&0\end{pmatrix} \quad \lambda_7 = \begin{pmatrix}0&0&0\\0&0&-i\\0&i&0\end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix}1&0&0\\0&1&0\\0&0&-2\end{pmatrix}
\end{aligned} \tag{2.1}$$

and are called the Gell-Mann matrices. They correspond to the Pauli matrices for SU(2).

The group SU(2) corresponds to the rotations in three-dimensional space. (There is actually not a one-to-one but a two-to-one mapping between SU(2) and the rotations because $\mathbf{U}$ and $-\mathbf{U}$ correspond to the same rotation.) The group SU(3) and other SU(n) are less easy to visualize because they do not correspond to something as simple as rotations. To understand a system with two or three quarks and antiquarks, a system of two spins can be studied which has four states $|\uparrow\uparrow\rangle$, $|\uparrow\downarrow\rangle$, $|\downarrow\uparrow\rangle$, and $|\downarrow\downarrow\rangle$. Three of them have spin 1, and one has spin 0, because there is nothing it can transform into.

In SU(3), there are three states and therefore $3 \times 3$ matrices. In Quantum Chronodynamics, the three states are the three colors. With

$$|r\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |g\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |b\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad |q\rangle = \begin{pmatrix} q_r \\ q_g \\ q_b \end{pmatrix} \qquad \mathbf{U} = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}$$

for one quark, starting from a quark in state $|r\rangle$ and applying $\mathbf{U}$ one gets a mixed state. Similar to the spin, where the different representations of SU(2) are spin 0, $\frac{1}{2}$, 1, $\frac{3}{2}$ and so on, there are the different representations of SU(3). For SU(2), there are matrix representations for every dimensionality, but this is not the case for SU(3). Three representations of SU(3) are of special interest.

The first representation is the representation of $3 \times 3$ matrices $\mathbf{U}$ that acts on a single quark and mixing up its different states. This representation is called 3 for three states, and can be written as $\mathbf{U}\,|q\rangle = |q'\rangle$. It is the transformation properties of the quarks.

There are also antiquarks, and their field operators are the complex conjugate of the field operators of the quarks such that $\mathbf{U}^*\,|q^*\rangle = |q'^*\rangle$. Therefore, there is a second representation which consists of the complex conjugates of the matrices of representation 3. It is called the complex conjugate representation and denoted by $\overline{3}$ (or sometimes by $3^*$). It is the transformation properties of the antiquarks.

Similar to taking two electron spins to give four states, two quarks can be taken to form a system with nine states. The corresponding representation is denoted by $3 \otimes 3$. This results, surprisingly, in an antiquark and a six-dimensional representation such that $3 \otimes 3 = \overline{3} + 6$. (There is a six-dimensional representation of SU(3), but it is not used here.) One can combine two quarks antisymmetrically which gives the $\overline{3}$ or symmetrically which gives the 6.

If one takes a quark and an antiquark leading to $3 \otimes \overline{3}$ this gives $1 + 8$ where the 1 is the singlet (a superposition of all colors and anticolors) which is completely invariant under color transformations and the 8 is the eight-dimensional representation of SU(3). When the color space is transformed the eight generators get mixed up.

Three quarks give $3 \otimes 3 \otimes 3$ which can be split up first into a $3 \otimes 3 = \overline{3} + 6$ and combined in a second step to $3 \otimes \overline{3} = 1 + 8$ or $6 \otimes 3 = 8 + 10$. Interesting is only the $1 + 8$ again and that one gets also a singlet plus the eight-dimensional representation of SU(3). The singlet is completely colorless and invariant under the SU(3) transformations. It is a red, a green and a blue quark antisymmetrized. The other combinations are red, red, green and so on. Important here is that one gets a singlet plus other stuff when combining a quark and an antiquark, and one gets a singlet plus other stuff when combining three quarks.

The gluon behaves as if it consists of a quark and antiquark with respect to the color symmetry but only the eight in $3 \otimes \overline{3} = 1 + 8$. In other words, gluons transform in the same way as the generators of the group. A gluon is not neutral.

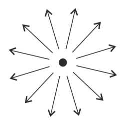
The postulates of Quantum Chromodynamics state first that it has SU(3) as the symmetry group. Further, a quark transforms as a 3 (triplet), the antiquark transforms as a $\overline{3}$ (antitriplet), and a gluon or gluon field transforms as an 8 which is a piece of a quark-antiquark. That is group theory as applied to Quantum Chromodynamics. Another postulate or better a dynamical output of the theory is that all real particles of nature, not quarks and gluons which are never seen singly, but all free particles found in the laboratories are always transforming under the 1. In other words, they are singlets.

There are only two fundamental ways of making singlets. One is to take a quark and an antiquark, the other is to take three quarks. In addition, one can take a singlet built from quark and antiquark and combine it with a singlet built from three quarks and so on, but they are all made out of the two kind of singlets. An object made out of three quarks is a baryon with half integer spin, and an object made out of a quark and an antiquark is a meson with integer spin.

There are the so-called glueballs made out of two or more gluons. If one takes two gluons giving 64 states and $8 \otimes 8 = 1 + 63$, there is again a singlet state. (The 64 states with two gluons is actually $8 \otimes 8 = 1+8+8+10+10+27$. This means that the $8\times 8$ matrix can be turned in a particular basis into a matrix with block structure in form of one $1\times 1$, two $8\times 8$, two $10\times 10$ and one $27\times 27$ blocks.) Thus, the spectrum of hadrons coming from quarks and gluons consists of baryons, mesons and glueballs. They are all color singlets. With three quarks or with a quark and an antiquark, one cannot make fractional electrical charge. The statement that electrical charge exists only as integer values in nature is equivalent to the statement that only SU(3) singlets (color singlets) and combinations thereof exist as free particles in nature. Color singlets are entangled states.

## 2.6 Interaction Between Quarks and Gluons

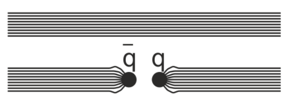
Gluons interact with quarks the same way as photons interact with electrons. In other words, gluons play the same role in Quantum Chromodynamics as photons play in Quantum Electrodynamics. Photons couple to the electric charge which is a conserved quantity. Thus, there must also be eight conserved quantities similar to electric charge, and each is the source of the particular gluon. The eight gluons correspond to fields which are similar to the electromagnetic field sourced by electrical charge. There are eight quantities namely the eight generators of SU(3) which are analogous to angular momentum in SU(2), and the things that radiate in the gluon field are the colors or the generators of the colors. Each of them can emit the right kind of gluon.



If one takes a colored object which is a quark, it is surrounded by a gluon field similar to the electric field of an electrically charged particle. An electric charge and anticharge attract each other. When bringing them close together, the energy gets lowered. The field energy in a neutral system is smaller than in a system with a net charge. The larger the charge of an electron, the more profitable it is to have a neutral system energetically, and the more expensive it is in terms of energy to have a system with a net charge. Thus, if the coupling constant of the electric charge would be much greater, low mass particles would be electrically neutral. If the charge is big enough, electrically charged particles would never have been detected. Consequently, if the coupling constant of Quantum Chromodynamics is big enough, one does not see single quarks in nature. The neutral state is the singlet state, and it is not possible to pull quarks apart into non-singlet states. Net amounts of color would cost a large amount of energy.

The dynamics in Quantum Chromodynamics is interesting because unlike the photon which is electrically neutral, the gluon has color. The fact that the photons are neutral means that they do not interact and that electromagnetic waves just pass each other. Electrodynamics is a linear theory. In Quantum Chromodynamics, the gluon as the analog of the photon is charged, and the analog of the electromagnetic wave does not only interact with other waves but parts of a gluon wave interact with other parts of the same wave. The field analogous to the electromagnetic field is much more interesting.


The equations related to the gluon field are very complicated, but the pictures are fairly easy. As an effect of the non-linearity, the lines of flux from a quark to an antiquark attract each other and create tubes. This is very different from electrodynamics and the lines of flux between two charged particles. There is an uniform field between the quark and antiquark. A flux tube (or fluxoid) does not get weaker when the distance between quark and antiquark get larger and larger, but the field energy increases. There is always a fixed number of lines independent of the distance, and the field is not diluting.




The effect is that the energy grows and grows, and the tube never gets thinner and thinner, when the quark and antiquark get separated until the system runs out of energy or the tube is broken into a pair of quark and antiquark. One meson, for example, can split into two or more mesons. Still there is no single quark. Any charged object with non-zero color, but not being a singlet, creates lines of flux going to infinity and therefore would need infinite energy. If one hits a meson, one of the particle in the quark-antiquark pair goes off, and the energy increases linearly until the kinetic energy is used up and it gets pulled back, or the fluxoid breaks and a quark-antiquark pair spontaneously gets created out of vacuum. If there is still enough kinetic energy, other quark-antiquark pairs may be created. Therefore, a jet of mesons is created when the quark is hit hard enough. Baryons exist because the flux lines can come together in threes.

## 2.7 Gauge Theories

Whenever there is a conserved quantity analogously to electric charge or the color which works as the source of a photon-like or gluon-like field, this is called a gauge theory. These theories are based on a symmetry because a conserved quantity like electric charge means a symmetry of some kind. Charged particles are the source of flux, and the flux lines must end at some charged particle. Thus, a charge cannot disappear because signals cannot travel faster than the speed of light, and lines of flux would have to disappear simultaneously. Removal of a charge means removal of the whole field also at infinity. Quantum Electrodynamics and Quantum Chromodynamics are two gauge theories. The Standard Model contains another three generators for the weak force in addition to the photon in Quantum Electrodynamics which is the theory of the electromagnetic force and the eight gluons in Quantum Chromodynamics which is the theory of the strong force.

The electromagnetic field is described by a four-vector $A_\mu$ with components $A_0$ and $\vec{A}$ where $A_0$ is the electrostatic potential and its gradient is the electric force with the $x$-component $F_x = e\partial A_0/\partial x$ and similar for $y$ and $z$. The existence of electromagnetic waves follows from Maxwell's equations. These electromagnetic waves have a direction and a polarization. In the figure the electric field is vertical and the magnetic field is horizontal. The gauge symmetry corresponding to the conservation of electrical charge is multiplication with a phase $\Psi \to e^{i\vartheta}\Psi$ which is called U(1), but this leaves the photon untouched because it is electrically neutral.

Every gauge field has sources. Baryon number, for example, could be the source of its own gauge field. This would create forces, and two neutrons would repel each other because they have the same baryon number, and a neutron and an antineutron would attract, but there is no such field for baryon number.

In Quantum Chromodynamics, the Maxwell-like field $A_{jk}$ is a matrix that transforms under SU(3). If $q_j$ is the component of a quark with color $j$, the symmetry operation is $U_{jk}q_k = q'_j$ and $U^*_{jk}q^*_k = q'^*_j$ for particle an antiparticle, and this operation is unitary. The fundamental representation are the colors red, green, blue, and the mathematical relation between the wave function of a particle and the wave function of an antiparticle is the complex conjugate. Unlike the photon which does not have electric charge, the gluons as the gauge bosons have color. Because of that, field lines get pulled into flux tubes.

## 2.8 Numbers and Orders of Magnitude in Atomic Physics

Coupling constants correspond to the probability of events. The question is, for example, what is the probability that an electron which is stopped suddenly emits a photon. The probability is pretty much $\alpha = e^2/(4\pi) \approx 1/137$ which is called the fine-structure constant. Thus, the answer to the question is that the probability is less than one percent. If one has the picture that a ray of photons is emitted when an electron hits a cathode-ray tube, the picture is wrong. The right picture is that, when a hundred-and-thirty-seven electrons hit the cathode-ray tube, in the average one of them emits a photon. In this sense, the electromagnetism is a weak force.

Energy, distance and time do not have to be thought of as different units. They are related with each other. Time and distance should not be distinguished in terms of units for fundamental physics because they have the same units when the speed of light $c$ is set to one. The unit of energy is the inverse of the unit of time with also Planck's constant $\hbar$ set to one. Thus, $1\,\mathrm{eV}^{-1} \approx 10^{-7}\,\mathrm{m}$. In other words, there is no need for different units for energy, space and time. The relation between energy and time is inverse such that big energies correspond to small time intervals and vice versa.

The diameter of a hydrogen atom is in the order of $10^{-10}\,\mathrm{m}$, and the transit time for light to cross a hydrogen atom is about $10^{-18}\,\mathrm{s}$. The time for an electron to orbit an atom is approximately $10^{-16}\,\mathrm{s}$ such that the speed of the electron is about one percent of the speed of light. The acceleration on the electron is not big and the orbit is quite large. This just means that the electromagnetic force on an electron is weak. The decay time for an excited hydrogen atom where the electron is just one state above the ground state is about $10^{-9}\,\mathrm{s}$. All these scales are related by powers of the fine-structure constant $\alpha$.

One can ask the same kind of questions for hadrons. The diameter is about $10^{-15}\,\mathrm{m}$, and the transit time for light is therefore about $10^{-23}\,\mathrm{s}$. The time for a quark to swing around in a proton is also in the order of $10^{-23}$ s. The typical decay time for an excited hadron is again about $10^{-23}$ s where it is assumed as an example that a proton gets hit, starts oscillating and emits a pion. The conclusion is that the analog of the fine-structure constant is close to 1, and therefore the probability for it to emit a gluon when stopped analogously to the electron by the cathode-ray tube is about 1. (The real number corresponding to the fine-structure constant is $\alpha_{\mathrm{QCD}} \approx 0.2$.) That is the reason why the strong interaction has been named strong interaction long before the development of Quantum Chromodynamics.

## 2.9 Weak Interaction

There are other processes called weak interactions. An example is the decay of a neutron $n \to e^- + p + \overline{\nu}$, and the lifetime is about 12 minutes. One of the reasons is that the mass of a neutron is only slightly bigger than the masses of the electron, proton and antineutrino. If the mass of a neutron would only be a bit smaller, no decay would happen at all because of energy conservation. Other examples of a weak decay are $\Pi^- \to e^- + \overline{\nu}$ and $\Pi^+ \to e^+ + \nu$ with lifetime in the order of ten nanoseconds. Here the energy available is not the reason why they decay so slowly. There must be something else that slows down the processes. The oldest known example of a weak interaction was the decay of the neutron which is called $\beta$-decay of the neutron. The other two shown examples are called the $\beta$-decay of the pion.

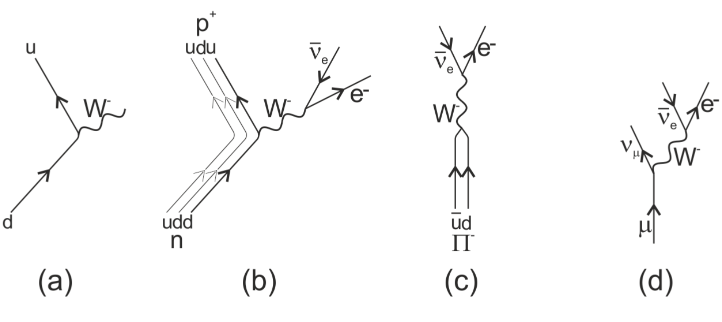
The list of quarks with the various possible color combinations such as $u_{\mathrm{r}}$ for up-quark with color red can be extended by an additional group of particles called leptons:

| | up | down | charm | strange | top | bottom |
|---|---|---|---|---|---|---|
| red | $u_{\mathrm{r}}$ | $d_{\mathrm{r}}$ | $c_{\mathrm{r}}$ | $s_{\mathrm{r}}$ | $t_{\mathrm{r}}$ | $b_{\mathrm{r}}$ |
| green | $u_{\mathrm{g}}$ | $d_{\mathrm{g}}$ | $c_{\mathrm{g}}$ | $s_{\mathrm{g}}$ | $t_{\mathrm{g}}$ | $b_{\mathrm{g}}$ |
| blue | $u_{\mathrm{b}}$ | $d_{\mathrm{b}}$ | $c_{\mathrm{b}}$ | $s_{\mathrm{b}}$ | $t_{\mathrm{b}}$ | $b_{\mathrm{b}}$ |
| leptons | $\nu_e$ | $e$ | $\nu_\mu$ | $\mu$ | $\nu_\tau$ | $\tau$ |

The strong interaction is based on the color symmetry which mixes up the colors and acts vertically in this table by connecting red, green, blue. It may change a red up-quark into a green up-park, but it does not mix up-quarks with down-quarks or charm-quarks and so on. The weak interaction acts horizontally and mixes up quarks with down quarks, charm with strange, and top with bottom. The weak interaction is based on SU(2), and this is also a gauge symmetry called flavor-symmetry.

The leptons are the electron, the muon, the tauon, and the corresponding neutrinos electron-neutrino, muon-neutrino, tauon-neutrino. This fourth row is not a color and these particles do not interact with gluons, but nevertheless it has also doublets as the quarks do, and the weak interaction acts similarly on these doublets. The difference between up and down is one unit of charge, and the difference between $\nu_e$ and e is also one unit of charge.

The gauge-bosons of the weak interaction are the W-bosons. Looking at one of the particles on the left side of a doublet and the antiparticle of the right side of the same doublet, there are the possible processes $\mathrm{e}^- + \overline{\nu}_e \leftrightarrow \mathrm{W}^- \leftrightarrow \mathrm{d} + \overline{\mathrm{u}}$ and $\mathrm{e}^+ + \nu_e \leftrightarrow \mathrm{W}^+ \leftrightarrow \overline{\mathrm{d}} + \mathrm{u}$. The W-bosons play the same role for weak interactions as the photons play for Quantum Electrodynamics and the guons for Quantum Chromodynamics. An electron $\mathrm{e}^-$ can become a neutrino $\nu_e$ by emitting a $\mathrm{W}^-$ or absorbing a $\mathrm{W}^+$.




Weak interactions do not change the color of quarks but can turn a down quark into an up quark with the same color as illustrate in the figure (a). This process is part of the $\beta$-decay of the neutron into a proton, an electron, and an electron-antineutrino shown in (b). A pion $\Pi^-$ can decay into an electron and an electron-antineutrino as in (c) but it is more probable that the pion decays into a muon and a muon-antineutrino and not into an electron and an electron-antineutrino. This are processes where quarks are involved, but there are also purely leptonic processes such as the decay of a muon into a muon-neutrino, an electron and an electron-antineutrino in (d). In general, electrons and their neutrinos can be replaced by muons and their neutrinos or tauons and their neutrinos.

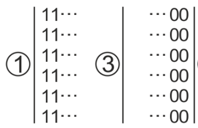
Weak interactions give very slow decay rates, and the analog of the fine-structure constant could be very small. However, this is not the reason because it is comparable to the one for electromagnetism, but the $\beta$-decay of a neutron is a very rare event because there is not enough energy in the neutron to create a real W-boson due to its mass (a hundred times the proton mass). Energy conservation can be violated for a very short time to allow virtual W-bosons because of the uncertainty principle. However, for the final state, energy and momentum must be conserved again, for example, by decaying into an electron and an antineutrino. This situation is similar to tunneling out of a potential without enough energy. Virtual W-bosons as all virtual particles cannot be observed. If one wants to observe phenomena in Quantum Mechanics which seem impossible such as tunneling or virtual particles, the seemingly impossible thing does no longer happen. To observe a virtual W-boson one needs to hit it with so much energy that it becomes a real one.

When a real W-boson decays, it produces very fast and therefore relativistic electrons. This leads to the question what particles are elementary. If a particle is hit with enough energy, many things can happen. If one hits, for example, an electron hard enough with a photon, many protons may come out.

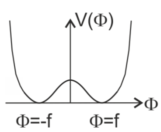
The Standard Model consisting of Quantum Electrodynamics and Quantum Chromodynamics together with the weak force is a gauge theory based on SU(3)×SU(2)×U(1). The weak interaction is covered by SU(2), and SU(2) has three generators. Thus, in addition to the two W-bosons, also a third particle called Z-boson is needed. It is its own antiparticle, while the $W^{\pm}$ are each others antiparticle.

## 2.10 Spontaneous Symmetry Breaking

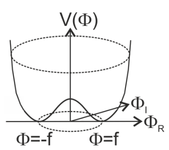
A lattice where each point can assume one of two states 0 or 1 and where antiparallel costs more energy than parallel will end up in a ground state where either all points have the value 0 or all points have the value 1. The system has a symmetry but it has more than one ground state. Explicit symmetry breaking is when one point needs less energy in one of the two states independent of the other points, and all other points in the lattice will also take on this value. In this case there is only one ground state.




All points in the same state is not always possible because two states may be forced by external constraints. If the system is forced along the line at ① in the figure to be in one state and along the line at ② to be in the other state, somewhere at ③ there must be a boundary where neighboring points have different values. This line is called the domain wall, and the effect is called spontaneous symmetry breaking. The difference between explicit and spontaneous symmetry breaking is the existence of domain walls in certain circumstances. In any case, the system assumes a state with lowest energy. If two external forces freeze different points in different states, the points of the whole lattice cannot assume the same ground state, and there must be domain walls which try to minimize the wrongly aligned points.




If there is a real, relativistic field $\Phi$ with Lagrangian $\mathcal{L} = \frac{1}{2}\partial_\mu\Phi\partial^\mu\Phi - V(\Phi)$ where the first term corresponds to the kinetic energy and the second term to a symmetric potential as shown in the figure, the system has two ground states $\Phi = \pm f$. Everywhere in space, the energy is therefore lowest for $\Phi = \pm f$, but it cannot jump from $\Phi = -f$ to $\Phi = f$ because it costs a large amount of energy. At high temperature, the random fluctuations may lead to plus and minus values, and when it cools down there is a random chance which ground state is assumed. Domain walls are created, and this is a case of spontaneous symmetry breaking despite the cost in energy.



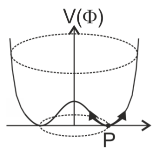

If the relativistic field is a complex field $\Phi = \Phi_\mathrm{R} + i\Phi_\mathrm{I}$ with the kinetic term in the Lagrangian $\partial_\mu\Phi^*\partial^\mu\Phi$, the field has a U(1) symmetry which is continuous. Replacing $\Phi$ by $\Phi' = e^{i\vartheta}\Phi$ and assuming that the potential also satisfies the U(1) symmetry gives a potential in the form $V(\Phi^*\Phi)$. The Lagrangian $\mathcal{L} = \partial_\mu\Phi^*\partial^\mu\Phi - V(\Phi^*\Phi)$ is then independent of rotations in the complex plane. If the potential has its minimum for

$\Phi = 0$ as in the case of $V(\Phi) = \Phi^*\Phi$, there is no spontaneous symmetry breaking. If, however, the potential has the form of a so-called Mexican hat because of $V(\Phi) = -a\Phi^*\Phi + b(\Phi^*\Phi)^2$ with $a, b > 0$, the minimum is not a point but lies on a circle in the complex plane with radius $f$ depending on $a$ and $b$. For small values of the field the term $-a\Phi^*\Phi$ dominates, and for large values of the field $+b(\Phi^*\Phi)^2$ dominates. The ground states corresponds to all the points on this circle with radius $f$. But even if it stays on this circle, there is a cost of energy to change from point to point due to the kinetic term $\partial_\mu\Phi^*\partial^\mu\Phi$ of the Lagrangian.

If some points in space are forced into $V(\Phi) = f$ and others into $V(\Phi) = if$, for example, there are no hard domain walls where $V(\Phi)$ jumps from $f$ to $if$, but the value can change gradually and smoothly along the shorter way between the two points on the circle. A transition like that is cheaper in terms of energy. Such a situation where the energy cost is just a gradient corresponds to a wave staying in this circle with lowest energy and to a massless particle.

When the energy of a wave $e^{ikx}$ which has a momentum $k$ goes to zero as $k$ goes to zero, that means that the quanta of this field has zero energy at rest. A particle with zero energy at rest is massless. This also means that the wavelength gets larger and larger. A term of the form $\frac{1}{2}m^2\,\Phi^*\Phi$ is called a mass term and it means that the particle has mass.

With the Lagrangian $\mathcal{L} = \partial\Phi\partial\Phi^* - V(\Phi^*\Phi)$ the field can be written as $\Phi = \rho e^{i\alpha}$, and the Lagrangian can be rewritten as $\mathcal{L} = (\partial\rho)^2 + \rho^2(\partial\alpha)^2 - V(\rho)$ because $\partial\Phi = (\partial\rho + i\rho\partial\alpha)e^{i\alpha}$ and $\partial\Phi^* = (\partial\rho - i\rho\partial\alpha)e^{-i\alpha}$. If $\rho$ does not change, it can be thought of as a constant $\rho = f$, and $\rho^2(\partial\alpha)^2 = f^2(\partial\alpha)^2 = (\partial\beta)^2$ where $\beta = f\alpha$. That is the characteristic of a field with no mass, and $\mathcal{L} = (\partial\beta)^2$ is the Lagrangian of a massless field with no potential energy at all. This means that if one is only allowed to vary the angle with fixed value $\rho = f$ then excitations of this type behave like fields whose quanta have no mass.




On the other hand, $\rho$ is an independent field. If the field gets moved from point $P$ in the figure to a higher value of the potential, it starts oscillating perpendicularly to the circle with radius $f$ back and forth about point $P$. If one takes the field everywhere and displace it in direction $\rho$, it starts oscillating, and that corresponds to a mass, because mass is energy associated with zero momentum and zero momentum means that the field does not vary from point to point. The frequency with which the system rocks back and forth due to the restoring force is the mass. A massive field oscillates when excited homogeneously. When displacing the field $\rho e^{i\alpha}$ on the other hand in direction $\alpha$, nothing much happens because there is no restoring force. Whenever one has a continuous symmetry, and that symmetry is spontaneously broken, there is always the possibility of changing the orientation gradually and one has massless excitations. Those particles are called Goldstone bosons. The corresponding Goldstone field is massless and has no restoring force by virtue of the symmetry. (They also exist in other parts of physics such as condensed matter physics.)

## 2.11 Gauge Invariance

Charges of various kinds are always carried by a complex field and the conservation law is associated with a symmetry. The symmetry for the electric charge is U(1) such that the Lagrangian is invariant under the transformation $\Phi \to e^{i\vartheta}\Phi$. The question is whether $\vartheta$ can vary from place to place. To see whether it is still a symmetry one can test whether it changes the Lagrangian. With $\Phi' = e^{i\vartheta}\Phi$ and $\Phi'^* = e^{-i\vartheta}\Phi^*$ where $\vartheta = \vartheta(x)$, one gets $\partial\Phi' = (\partial\Phi + i\Phi\,\partial\vartheta/\partial x)e^{i\vartheta}$ and $\partial\Phi'^* = (\partial\Phi^* - i\Phi^*\,\partial\vartheta/\partial x)e^{-i\vartheta}$ such that

$$\partial\Phi'\partial\Phi'^* = \partial\Phi\partial\Phi^* + i[\Phi\partial\Phi^* - \Phi^*\partial\Phi]\frac{\partial\vartheta(x)}{\partial x} + \Phi\Phi^*\left(\frac{\partial\vartheta(x)}{\partial x}\right)^2$$

because $e^{i\vartheta}$ and $e^{-i\vartheta}$ cancel, but this result is not a symmetry because $\partial\Phi'\partial\Phi'^* \neq \partial\Phi\partial\Phi^*$. If the potential energy only depend on $\Phi^*\Phi$ the potential term in the Lagrangian is symmetric, but the kinetic term is not because of the derivative of $\vartheta(x)$ not being zero. What is actually happening here is that a little bit of gradient energy is added by making $\vartheta$ vary from place to place.

Being a gauge symmetry implies that the symmetry parameter ($\vartheta$ in the case of the electric charge) can vary from place to place, and as an experimental fact all interactions of nature including photons, gluons, Z- and W-bosons are based on a gauge symmetry. Thus, there must be a way to restore symmetry, and indeed there is one, but one has to introduce more fields as gauge boson fields. For electric charge, it is the electromagnetic field defined via a vector potential through the four-vector $A_\mu$. It transforms as $A_\mu \to A'_\mu = A_\mu - \partial_\mu \vartheta$. As a reminder, $A_0$ is the electrostatic potential whose gradient in space is the electric field, and the curl of $A_i$ is the magnetic field.

This needs a new definition of a Lagrangian where the derivative is replaced by the so-called covariant derivative (it is not the covariant derivative of General Relativity) defined as $D_\mu \Phi = \partial_\mu \Phi + iA_\mu \Phi$ and $D_\mu \Phi^* = \partial_\mu \Phi^* - iA_\mu \Phi^*$. The kinetic term in the Lagrangian $\partial\Phi\partial\Phi^*$ becomes $D\Phi D\Phi^*$. Instead of looking at the whole Lagrangian, it is sufficient to compare the covariant derivative of $\Phi'$ with the one of $\Phi$. The covariant derivative of $\Phi'$ is

$$D\Phi' = ([\partial\Phi + i\Phi\partial\vartheta] + i(A - \partial\vartheta)\Phi)\, e^{i\vartheta} = (\partial\Phi + iA\Phi)\, e^{i\vartheta} = D\Phi e^{i\vartheta}$$

where the terms $\pm i\Phi\partial\vartheta$ cancel. Thus, $D\Phi' = D\Phi e^{i\vartheta}$ and $D\Phi'^* = D\Phi^* e^{-i\vartheta}$ and the term $D\Phi D\Phi^*$ in the Lagrangian is gauge invariant, and this is due to the introduction of the field $A$ with the transformation property $A_\mu \to A'_\mu = A_\mu - \partial_\mu \vartheta$.

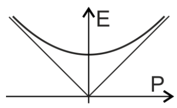
The full Lagrangian of a simple gauge theory involving the interaction of electromagnetism with a charged scalar field is $\mathcal{L} = D\Phi^* D\Phi - V(\Phi^*\Phi) + F^2$ with $F^2 = F_{\mu\nu}F^{\mu\nu}$ because the electromagnetic field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is also gauge invariant since the transform $A_\mu \to A'_\mu = A_\mu - \partial_\mu \vartheta$ ensures that the cross-terms $\partial_\mu \partial_\nu \vartheta$ cancel. The electromagnetic field $A$ is coupled to a field $\Phi$ whose quanta are charged particles, and the initially global symmetry has become a local gauge symmetry.

One cannot add potentials to the Lagrangian which are not a function of $\Phi^*\Phi$, but one can also not add terms like $\frac{1}{2}m^2 A_\mu A^\mu$ which would give the photon mass. This term would not be gauge invariant because $\frac{1}{2}m^2 A_\mu A^\mu \neq \frac{1}{2}m^2 A'_\mu A'^\mu$. Thus, if one believes in gauge invariance, and every physicist does, the photon cannot have a mass.

## 2.12 Gauge Particles and Mass

Spontaneous breaking the symmetry U(1) through a Mexican hat potential would give the photon a mass as it does for the Z-boson and the W-bosons, and it would also miraculously remove the Goldstone bosons. (The Goldstone boson is a feature of the theory without vector potentials and without gauge invariance. The energy of the Goldstone boson was exactly the energy for making these variable phases. The introduction of the vector potential shows a way for making the variable phases without a cost in energy because the Lagrangian is invariant under these gradual changes of phase.) This mechanism is called the Higgs phenomenon. Spontaneous symmetry breaking can give a particle mass where one cannot add a mass term to the Lagrangian. However, the symmetry of the photon is not spontaneously broken, and the photon does therefore also not get mass through the Higgs mechanism.

The photon does not have a mass, but if the U(1) symmetry as the gauge symmetry for electric charge would be spontaneously broken, the Higgs mechanism would give it a mass. (To show this mechanism for the photon is much easier than for the Z-boson and the W-boson which really get their mass through this mechanism.) There is actually a situation in nature where the photon gets a mass and the U(1) symmetry is spontaneously broken because the photons in a superconductor propagate with a mass due to the Higgs phenomenon. As soon as they leave the superconductor, they are again massless.




However, the photon in a prisma does not move slower than with the speed of light due to a mass. Its energy and therefore its mass is still zero when momentum is zero. The energy of a particle with mass is $E = \sqrt{p^2 + m^2}$ or $E^2 = p^2 + m^2$ for $c = 1$. For a massless particle, the relation between $E$ and $\pm p$ is either of the two lines through the origin in the figure. For a particle with mass, the relation is a hyperbola, and the lowest point is its mass. If one shifts the homogeneous field a bit and let it go, it starts to oscillate, and particles with mass at rest oscillate because they have mass. Oscillations of a field when it is homogeneous or, in other words, when it has infinite wavelength, this is what is called mass. When light enters glass, the slope of the line through the origin gets smaller, but it is still a straight line through the origin, and all photons move with the same speed. One spots a massless object when there is a flat direction of a potential. If the potential is not flat, the field starts to oscillate when shifted away from equilibrium, and this is called mass. A field with a Mexican hat potential has two distinct quanta associated with it. The massless Goldstone boson corresponds to the flat motion changing $\alpha$ for the fixed value $\rho = f$, and the Higgs boson with mass corresponds to the oscillation around $\rho = f$ for a fixed value $\alpha$. However, there is no massless Goldstone boson in nature, because the Goldstone boson got eaten by the gauge boson resulting in giving the Higgs boson a mass as described below in more details.

There are particles in the Standard Model which for one mathematical reason or another cannot have mass unless the symmetry is spontaneously broken. There can be others that do not require this mechanism. They are the ones one cannot see in the laboratories, because one can only see the light particles. There are strong reasons to believe that there are additional particles in nature. Dark matter particles could be a candidate. It is assumed that they are a thousand times heavier than protons. Those particles with large mass do not require spontaneous symmetry breaking to get a mass.

## 2.13 Higgs Fields and Gauge Bosons

The first class of particles which cannot have a mass without spontaneous symmetry breaking are the gauge bosons. The starting point for the Higgs mechanism is a boson field $\Phi$ which is a complex field $\Phi = \Phi_\mathrm{R} + i\Phi_\mathrm{I} = \rho e^{i\alpha}$ with $\Phi^* = \Phi_\mathrm{R} - i\Phi_\mathrm{I} = \rho e^{-i\alpha}$ where $\rho$ and $\alpha$ are functions of position and where $\Phi_\mathrm{R}$ and $\Phi_\mathrm{I}$ respectively $\rho$ and $\alpha$ are two separate fields. For such fields, multiplication with a phase does not change $\partial\Phi^*\partial\Phi$. With a Mexican hat potential the Higgs boson $\rho$ gets a mass, the Goldstone boson $\alpha$ does not. The Lagrangian of $\Phi$ is $\mathcal{L} = \partial\Phi^*\partial\Phi - V(\rho)$ with $\partial\Phi^*\partial\Phi = (\partial\Phi_\mathrm{R})^2 + (\partial\Phi_\mathrm{I})^2 = (\partial\rho)^2 + \rho^2(\partial\alpha)^2$ where the potential $V$ has a U(1) symmetry.

If the potential is $V(\rho) = \rho^2$ it can be written as $\Phi_\mathrm{R}^2 + \Phi_\mathrm{I}^2$ and the Lagrangian becomes $\mathcal{L} = (\partial\Phi_\mathrm{R})^2 + (\partial\Phi_\mathrm{I})^2 - \Phi_\mathrm{R}^2 - \Phi_\mathrm{I}^2$ such that they become two independent fields $(\partial\Phi_\mathrm{R})^2 - \Phi_\mathrm{R}^2$ and $(\partial\Phi_\mathrm{I})^2 - \Phi_\mathrm{I}^2$. The potential has the form of a parabola rotated around the axis, and both independent field get a mass.

With a Mexican hat potential instead, the form $\rho e^{i\alpha}$ is more convenient than $\Phi_\mathrm{R} + i\Phi_\mathrm{I}$. The energy has its minimum for $\rho = f$ independent of $\alpha$, but one value of $\alpha$ must have been assumed and the symmetry is therefore broken as discussed above. If the field $\rho$ is only slightly displaced away from $f$, it can be written as $\rho = f + H$. Thus, the Higgs field $H$ is just a displacement in radial direction away from equilibrium. For low-energy excitation $\rho \approx f$ the Lagrangian becomes $\mathcal{L} = f^2(\partial\alpha)^2$. There is no potential energy associated with $\alpha$ because one does not gain any energy when going around on the circle $\rho = f$. By rescaling the field to $\beta = f\alpha$ the Lagrangian becomes $(\partial\beta)^2$, and there is no mass term because it does not cost any energy to displace the field homogeneously everywhere. Thus, $\alpha$ (or $\beta$) is a massless field. It is the Goldstone boson. (Goldstone bosons are the consequences of the spontaneous symmetry breaking of a continuous symmetry.)

The other side of spontaneous symmetry breaking which also has to do with varying the phase of a field from point to point is the gauge invariance. The electromagnetic field with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ which is gauge invariant has the Lagrangian $F^2 = F_{\mu\nu}F^{\mu\nu}$. As shown above, the full Lagrangian is $\mathcal{L} = D\Phi^* D\Phi - V(\Phi^*\Phi) + F^2$ for a simple gauge theory involving the interaction of electromagnetism with a charged scalar field $\Phi$. The photon does not have a mass because $F^2$ is only a function of derivatives of the vector potential and therefore shifting the vector potential everywhere simultaneously does not change the energy. Thus, there are the Goldstone boson and the photon as the two massless particles in this theory. A mass term for the photon would be proportional to $A^2$, but would not be gauge invariant.

The gauge invariance disallowing a term $A^2$ in the Lagrangian prohibits the photon from having mass. The term $D\Phi^* D\Phi = (\partial\Phi^* - iA\Phi^*)(\partial\Phi + iA\Phi)$ on the other side contains a term $A^2\Phi^*\Phi = f^2A^2$ if $\Phi$ is stuck in the ground state $\Phi = f$ for the Mexican hat potential $V$. The term $f^2A^2$ looks like a mass term for the photon because $f^2$ is just a numerical factor. Thus, when the symmetry is spontaneously broken and when one studies the theory in the neighborhood of the minimum $\Phi = f$, the effect is that one of the terms in the Lagrangian is a mass term for the photon. (Real photons do not have mass, but photons in this theory with the Mexican hat potential for all practical purposes would have a mass.)

At the same time the Goldstone boson disappears. It is a slowly varying value of $\alpha$ which is a field and can vary in space. One can make $\alpha$ also varying in space by adding $\vartheta(x)$. This does not make a Goldstone boson because this is a gauge symmetry, and a symmetry means that one does not change the energy. This is not a real change because the system stays in the ground state. The Goldstone boson becomes a gauge transformation. The Higgs field $\rho$ can oscillate and gets therefore also a mass. That is the Higgs mechanism. (This, by the way, is also the physics of superconductors. In a superconductor there are charged bosons which are loosely bound pairs of electrons called Cooper pairs. The charged boson field is shifted away from the origin by some complicated dynamics in the superconductor. It creates a condensate for these charged pairs. The field $\Phi$ is in this case the field of the charged Cooper pairs. The consequence is that the photon in a superconductor behaves as if it has mass.) The Higgs mechanism does not give the photon an mass but gives the Z-boson and the W-bosons their mass. The mathematics is more difficult because the symmetry group is not just U(1) but SU(2) but it is similar to the calculations above. This solves to problem that the Z-boson and the W-bosons have a mass despite the fact that gauge bosons cannot have a mass term because such a term would not be gauge invariant.

The Mexican hat potential is steep in angular direction and it needs much energy to excite it out of the ground-state. The field $\rho e^{i\alpha}$ becomes $fe^{i\alpha}$ when it is stuck at the ground-state, and the covariant derivative is $D\Phi = \partial\Phi + iA\Phi = i(\partial\alpha + A)fe^{i\alpha}$ such that $D\Phi^* D\Phi = f^2(\partial\alpha + A)^2$. The transformation $A \to A' = A + \partial\alpha$ is just a gauge transformation. Thus, replacing $A$ with $A'$ does not change $F_{\mu\nu}$ because $F_{\mu\nu}$ is gauge invariant. The term $f^2 A'^2$ in the Lagrangian does not have derivatives in it. This is the mass for the gauge boson. This shows how the shift of the field from zero to $f$ gives a particle a mass which otherwise could not have a mass. The vacuum shifted to a new spontaneously broken vacuum, and the excitations of the electromagnetic field in this shifted vacuum have a mass. In addition, the field $\alpha$ has disappeared, and with it also the Goldstone boson has disappeared.

So far the oscillations in the angular direction have been ignored, but with enough energy the Higgs field can be excited out of the equilibrium in the ground-state such that it starts oscillating in the angular direction. Those oscillations are also quanta and they are massive quanta because they have a lot of energy. They are the Higgs bosons which has a mass.

It looks like the angular degree of freedom has disappeared but it did not disappear. A massless photon has two polarization states which are orthogonal to the direction of motion. The photon moves with the speed of light and can never be brought to rest. If the photon could be brought to rest, it could be accelerated in the direction of its polarization. Once at rest there is no distinction anymore between the three directions. Thus, if a photon would have mass there would be three directions of polarization. The Goldstone boson becomes the missing third direction of polarization.

## 2.14 The Higgs Field and Fermions

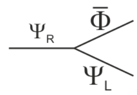
Reflection as mirroring symmetry is fulfilled in Quantum Electrodynamics and Quantum Chromodynamics, but it is violated by the weak interaction. Particles have a handedness also called helicity. The spin together with the direction of motion of an electron allows to distinguish between left-handed and right-handed electrons. It is an experimental fact that electrons in the $\beta$-decay of a neutron are always left-handed. In general, it is always a left-handed electron that interacts with the W-boson. It is, however, not possible that only left-handed and no right-handed electrons exist in nature because one can bring a left-handed electron to rest and accelerate it in the opposite direction to turn it into a right-handed electron. As a consequence it is not possible to define the handedness of an electron at rest. The notion of handedness is most useful for high-energy particles and for massless particles. The helicity of a massless particle cannot change because it cannot be brought to rest.

The Dirac equation has four components, two for positive and negative energy and two for the two directions of the spin. In the form $i(\partial\Psi/\partial t + \alpha_j \partial\Psi/\partial x_j) = m\beta\Psi$, the left side of the equation contains only derivatives of the field and does therefore not lead to a mass. The mass-term couples the left-handed component $\Psi_\mathrm{L}$ and right-handed component $\Psi_\mathrm{R}$ of $\Psi$ because $\beta$ switches left-handedness and right-handedness.

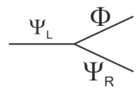
One can try to understand the asymmetry of the $\beta$-decay by the physically incorrect assumption that only left-handed electrons have electric charge but not right-handed electrons. Further, a charged boson field $\Phi$ is assumed together with the transformations $\Psi_\mathrm{L} \to e^{i\vartheta}\Psi_\mathrm{L}$, $\Psi_\mathrm{R} \to \Psi_\mathrm{R}$ and $\Phi \to e^{i\vartheta}\Phi$. When the mass $m$ is replaced by a numerical factor $g$ and the field $\Phi$, the Dirac equations become

$$i\left(\frac{\partial\Psi_\mathrm{R}}{\partial t} + \alpha_j \frac{\partial\Psi_\mathrm{R}}{\partial x_j}\right) = g\Phi^*\Psi_\mathrm{L} \qquad\qquad i\left(\frac{\partial\Psi_\mathrm{L}}{\partial t} - \alpha_j \frac{\partial\Psi_\mathrm{L}}{\partial x_j}\right) = g\Psi_\mathrm{R}\Phi$$

where $g$ is called the coupling constant. These two equations are consistent with the U(1) symmetry, and therefore they also do not violate charge conservation.




The two new processes possible with these two equations turn a right-handed particle into a left-handed particle or vice versa but only at the cost of emitting a charged particle to compensate for the charge. The amplitude is the coupling constant $g$ which is called the Yukawa coupling, and it is different for different fermions.




The Higgs field is a charge-carrying field. Thus, it is further assumed that $\Phi$ is a Higgs field $\Phi = \rho e^{i\alpha}$ with magnitude $f$, and the right sides of the two above Dirac equations become $g\Phi^*\Psi_\mathrm{L} \approx gf\Psi_\mathrm{L}$ and $g\Psi_\mathrm{R}\Phi \approx gf\Psi_\mathrm{R}$. The term $gf$ plays the same role as the mass $m$ in the two equation above. The mass here is just the coupling between left-handedness and right-handedness. Thus, the Higgs phenomenon has not only the role of giving mass to the Z-boson and the W-bosons, but it gives also mass to the fermions.

There are different fermions, and this means that there must also be different values for $g$ for the different fermions. These values are called Yukawa couplings. The value for electrons is $g_e \sim 10^{-5}$ and for top quarks $g_t \sim 1$. There is a whole spectrum of Yukawa couplings nobody understands. This procedure of associating such coupling constants to types of fermions sounds like fitting a collection of numbers to the masses of fermions, but these numbers might be experimentally confirmed. The term $f$ is the frozen value of the Higgs field, but the Higgs field is not frozen since it can shift around. Hit with enough energy it starts oscillating. The term $gf\Psi_\mathrm{L}$ should be written as $g(f-H)\Psi_\mathrm{L}$, and this term shows that a right-handed fermion can turn into a left-handed fermion and a Higgs particle. The Higgs boson has not been detected yet, but it is important to study the different processes by which a Higgs boson is produced when it has been detected to see whether the ratios are in agreement with these Yukawa couplings. As a consequence the Higgs boson is supposed to decay preferably into heavier fermions.

In the Standard Model of Particle Physics there is only one Higgs boson, but in other theories can be more than one. (The supersymmetric theory, for example, has two Higgs bosons.) In the Standard Model all massive particles get their mass from spontaneous symmetry breaking and the Higgs mechanism. If $f$ would be zero they would all be massless possibly with the exception of the Higgs boson itself.

One question is why are there no particles with masses much larger than $f$. There may be such particles but they have not been detected yet because the energies possible today are limited roughly at the level $f$ which is about 250 GeV. Dark matter particles are assumed to have a mass a factor of ten larger, and it may be the first particle whose mass does not come from the spontaneous symmetry breaking of the Higgs mechanism.

## 2.15 Neutrinos

Neutrinos have a very small mass, and they move within experimental error with the speed of light. This means that it is possible that they have only a single handedness. In the Standard Model they are massless and exist only left-handed. Generalizations of the theory have right-handed neutrinos, but they have typically a very large mass. It is known that the neutrinos have a very small mass because of neutrino oscillations. As shown above a mass term for fermions turns left-handed particles into right-handed particles. There are left-handed neutrinos and right-handed antineutrinos, and the mass term turn left-handed neutrinos into right-handed antineutrinos. The mass term cannot turn a left-handed electron into a right-handed positron because this process violates charge conservation. The neutrino has no charge. A particle which gets its mass by mixing with its antiparticle is called a Majorana particle.

For neutrinos, there are two views. One is that there are left-handed neutrinos and right-handed antineutrinos, and the other view is that there are left-handed and right-handed neutrinos where the neutrino is its own antiparticle. The neutrinos can mix together with antineutrinos and form a massive Majorana particle which is its own antiparticle because it is a superposition of a particle and an antiparticle. The neutrino is charged with respect to the weak interaction. The neutrino and antineutrino interact differently with the Z-boson and have therefore opposite charge with respect to the weak force.

## 2.16 Chirality and the Mass of Fermions

The Dirac equation (1.11) can be written as $i\partial_t\Psi + i\alpha_j\partial_j\Psi - \beta m\Psi = 0$. To get the Lagrangian, one can multiply $\Psi^\dagger$ with the left side of this equation. This is not the form in which the Dirac equation and the corresponding Lagrangian are usually written. With the definitions $\bar{\Psi} = \Psi^\dagger \beta$ (or equivalently $\Psi^\dagger = \bar{\Psi}\beta$ because $\beta^2 = \mathbf{I}$), $\gamma^0 = \beta$ and $\gamma^j = \beta\alpha_j$, the Dirac equation and the Lagrangian for the Dirac field become

$$(i\gamma^\mu \partial_\mu - m)\,\Psi = 0 \qquad\qquad \mathcal{L} = \bar{\Psi}\,(i\gamma^\mu \partial_\mu - m)\,\Psi \tag{2.2}$$

which gives them a much more elegant form. The matrices $\gamma^\mu$ are also called Dirac matrices. Note that $\bar{\Psi}\Psi$ is a scalar and $\bar{\Psi}\gamma^\mu\Psi$ is a 4-vector. There is also an additional $\gamma$ matrix $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$ with $(\gamma^5)^2 = \mathbf{I}$. It has two eigenvalues $-1$ and two eigenvalues $+1$ corresponding to the left-handedness and right-handedness, respectively, where handedness here means chirality. (Helicity and chirality are not the same, but the difference has not been explained.) The term $\bar{\Psi}\gamma^\mu\partial_\mu\Psi$ in (2.2) does not multiply left-handed times right-handed, but $\bar{\Psi}\Psi$ does because it is $\Psi_\mathrm{L}^\dagger\Psi_\mathrm{R} + \Psi_\mathrm{R}^\dagger\Psi_\mathrm{L}$. Thus, the mass term mixes the left-handed and right-handed particles. A massless fermion has a definite handedness, but the mass term flips the handedness back and forth.

The emission and absorption of W-bosons is restricted to left-handed fermions. (The Z-boson is different but has also different coupling to left-handed and right-handed fermions.) Thus, the W-bosons couple only thought the left-handed degrees of freedom such that the left-handed fermions are charged with respect to the weak interaction symmetry SU(2) and the right-handed fermions are uncharged. Therefore a term like $\Psi_\mathrm{L}^\dagger\Psi_\mathrm{R}$ which changes a charged particle into an uncharged particle is not allowed in the Lagrangian because it violates charge conservation.

The Higgs boson is also charged with respect to SU(2) of weak interactions because it is a doublet and also transforms under SU(2). To restore charge conservation the term $\Psi_\mathrm{L}^\dagger\Psi_\mathrm{R}$ has to be replaced by $\Psi_\mathrm{L}^\dagger\Phi\Psi_\mathrm{R}$ and the term $\Psi_\mathrm{R}^\dagger\Psi_\mathrm{L}$ by $\Psi_\mathrm{R}^\dagger\Psi_\mathrm{L}\Phi^\dagger$ where $\Phi$ is the Higgs field. The Feynman diagram corresponding to $\Psi_\mathrm{L}^\dagger\Phi\Psi_\mathrm{R}$ has an incoming left-handed fermion and an outgoing right-handed fermion together with a Higgs boson such that charge conservation is guaranteed.

However, the Higgs field has the Mexican hat potential such that $\Phi = f + H$ is a number and a fluctuation which is usually called the Higgs field. The terms $\Psi_\mathrm{L}^\dagger\Phi\Psi_\mathrm{R} + \Psi_\mathrm{R}^\dagger\Psi_\mathrm{L}\Phi^\dagger$ becomes

$$g_\mathrm{Y}(\Psi_\mathrm{L}^\dagger\Phi\Psi_\mathrm{R} + \Psi_\mathrm{R}^\dagger\Psi_\mathrm{L}\Phi^\dagger) = g_\mathrm{Y} f(\Psi_\mathrm{L}^\dagger\Psi_\mathrm{R} + \Psi_\mathrm{R}^\dagger\Psi_\mathrm{L}) + g_\mathrm{Y} H(\Psi_\mathrm{L}^\dagger\Psi_\mathrm{R} + \Psi_\mathrm{R}^\dagger\Psi_\mathrm{L})$$

written including the Yukawa coupling $g_\mathrm{Y}$. The first term $g_\mathrm{Y} f(\Psi_\mathrm{L}^\dagger\Psi_\mathrm{R} + \Psi_\mathrm{R}^\dagger\Psi_\mathrm{L})$ allows to determine the Yukawa couplings for the different fermions. The second term $g_\mathrm{Y} H\Psi^\dagger\Psi$ corresponds to processes such as an left-handed electron coming in and a right-handed electron plus a Higgs boson going out or a Higgs boson decaying into an electron and a positron. Higgs bosons prefer to decay into heavier fermions given enough energy, and this is reflected in the Yukawa couplings depending on the kind of fermions.

## 2.17 The Dynamics of the Higgs Boson

The dynamical behavior of the Higgs boson is basically fully determined by the Mexican hat potential. The minimum indicates how much the field is shifted and what the value $f$ is. Around zero the potential is more or less an upside-down parabola $-\mu^2/2\Phi^2$. With a plus sign this term would correspond to a mass term, but with a negative sign it is something like an imaginary mass. The point at the origin is not stable, and this term alone in the potential would make a vacuum with minimal energy impossible. Adding a positive term to turn the potential around would make $V = -\frac{1}{2}\mu^2\Phi^2 + \frac{1}{4}\lambda\Phi^4$. In order to keep the symmetry of the potential odd powers of $\Phi$ are not allowed, but higher even powers could be added. To get the minimum, $V$ is differentiated with respect to $\Phi$ resulting in $-\mu^2\Phi + \lambda\Phi^3 = 0$ and $f^2 = \mu^2/\lambda$. Thus, $f = \mu/\sqrt{\lambda}$ controls the mass of all known particles including the Higgs boson.

The question is why are all masses of known particles defined by $f$. If a fermion would exist whose left-handed and the right-handed version both carry charge of the weak interaction and react with the weak interaction the same way, this particle could have a mass term without the need of spontaneous symmetry breaking resulting in a mass which is not controlled by $f$. Particles which do not have weak interactions at all could have a mass.

The general thinking is that the natural mass scale for particle physics is enormous and much higher than the masses of any ordinary elementary particles. The only small mass parameter in nature is $f$. Those particles which have to have mass proportional to $f$ are the ones which would be massless without the shift $f$ of the Higgs field. The reason why all known particles get their mass this way is because the other particles have mass which is much larger and have therefore not been detected yet, and the question is why is $f$ so small (in the order of 100 GeV). The next question is compared to what is $f$ small.

## 2.18 Unification Scale Including Gravity

For a description of elementary particles one begins with a Lagrangian, and a Lagrangian has parameters such as mass and charges, but these quantities are rarely the measured quantities. The measured quantities are the product of all kinds of interactions which take place between the high-frequency modes of a system and the low-frequence modes of a system which change the values of the parameters that one associates with experimental observation.

An example is electric charge. The normally observed charge of an object is defined by the Coulomb law at large separation. If one asymptotically separates them, the potential energy between two electrons is $e^2/r$ and the force is $e^2/r^2$ (attractive for opposite charges and repulsive for like charges). In a material or even in the air the measured value for electric charge is always a little bit less than the one in physics tables which is only true in vacuum.

In an electric conductor a charged object is surrounded by a cloud of opposite charge which neutralizes it. Thus, between two charged objects in an electric conductor there is not force at all. If the charged objects of opposite charge are small compared to the cloud around them, and if one brings them closer together than the screening cloud, the two charged objects feel each others raw basic charge. The charge $e$ is therefore a function $e(r)$ of distance $r$. This is called the running charge.

A dielectric has electrons which are bound to the atoms but they are free to shift a little bit in an electric field. If one puts a negatively charged object into a dielectric, the electrons bound in the atom shifts a little bit away from this object. This does not screen completely as in the case of the electric conductor but only a fraction of the charge. The fraction is controlled by the dielectric constant. The electric field of this charged object is diminished. Also in this case there is a running charge $e(r)$ depending on $r$.

Exactly this phenomenon happens in Quantum Field Theory. In Quantum Electrodynamics, for example, the vacuum is full of electron-positron pairs. There are Feynman diagrams where such pairs are created by photons or even without photons as electron-positron loops. If an electromagnetic field comes along, it shifts the electrons one way and the positrons the other way. The virtual electron-positron pair becomes polarized. The vacuum is basically a dielectric. Thus, a positive charge put into the vacuum is screened by a little bit of negative charge, and a negative charge put into the vacuum is screened by a little bit of positive charge.

The charge that is felt at a large distance in all three cases is not the same as the numerical value of the charge one puts into the Lagrangian. If one moves two small charged objects closer and closer, one can get closer than the screening cloud. There is one difference because there is no set distance that the electron and the positron tend to be separated by. The right statement is that the separation of the electron-positron pair depends on their energy. The higher the energy of the virtual pair the closer they will be. Thus, there will be electrically neutral pairs with lower energy and larger size and others with higher energy and smaller size. If one brings two charged objects together, then one brings them first close enough that they are within the distance of the low-energy virtual pairs, and the low-energy pairs will no longer be able to screen the charge and the charge of the object will look a little bit bigger than when they were far apart. If one brings the two charged objects closer, also the high-energy pairs will no longer screen, and the function $e(r)$ keeps increasing gradually and does not stop at some small distance because there are virtual pairs of every possible energy in the vacuum.

When scattering electrons with low energy, they only get within a certain distance with each other. The higher the scattering energy the closer the electrons can come. Thus, for calculations one has to use the the running charge, and electrons seem to have a higher charge when scattered with high energy than when scattered with low energy.

This leads to the question what the fundamental charge is. The charge at long distance is affected by all the screening taking place. If one gets closer and closer one somehow peels away more and more of the screening cloud where the bare charge manifests itself. One starts to see the high-precision measurable effect of the running charge around the Compton wavelength of the electron which is about $10^{-11}$ cm. It would keep going on if the physics did not change somehow at very short distances. (At Planck length physics will certainly change.)

For the strong interactions things become more complicated. The force law between quarks turns the field lines into tube-like structures because of the interaction between gluons. As one separates quarks, the force grows linearly with the distance between them. Unlike in Quantum Electrodynamics where the energy grows with $e^2/r$ the energy grows with $e^2 \cdot r^2$ with $e$ corresponding to the charge of the strong force. The energy grows with separation but at very short distance it reverts back to $e^2/r$. The running charge $e(r)$ of the weak interactions is somewhere between the running charge of Quantum Electrodynamics and Quantum Chromodynamics.

If one plots the three running charges $e(r)$ (usually for $1/r$) the three curves meat at about $10^{16}$ GeV and it looks as if there is some kind of unification going on there. If this is correct the fundamental length scale of the Standard Model of Particle Physics is up at this level because the inputs seem to be simplest and the three coupling constants are the same. This energy also happens to be more or less the place if one plots the strength of the gravitational forces. They are very weak at large distances but they get stronger and stronger at small distances not only because of the usual $1/r^2$ law but above and beyond that. The plots do not cross exactly at the same place, but they cross within a couple of decades of energy of each other. If there is a unification scale this looks as if it is not far from the Planck scale where gravity becomes comparable with the other forces.

At small scale gravity is not weaker than the other forces where the law becomes $1/r^4$ and is no longer $1/r^2$. Newton's law $m_1 m_2 G/r^2$ with the product of the two masses $m_1$ and $m_2$ times Newton's constant $G$ in the numerator is a non-relativistic formula which is valid for particles at rest. It is energy which gravitates and not mass. The relativistic formula is $F = (E_1 E_2 G)/(c^4 r^2)$ where $E_1$ and $E_2$ are the energies of the two objects. In Quantum Mechanics bringing two particles very close together (much closer than their Compton wavelength) has the consequence that the momentum is very uncertain because of the uncertainty principle or, in other words, is very large in the order $p \sim \hbar/r$. These particles become highly relativistic with $E_1 = E_2 = E = pc$ and the gravitational force becomes

$$F = \frac{E_1 E_2}{c^4 r^2} G = \frac{\hbar^2 c^2}{c^4 r^4} G = \frac{\hbar^2}{c^2 r^4} G$$

such that the law is for very short distances $1/r^4$ .

The question is at what length scale $r$ is the gravitational force in the same order of magnitude as the electromagnetic force. Using the energy instead of the force (and with setting the electric charge $e = 1$ for simplicity) gives

$$\frac{e^2}{r} \sim \frac{\hbar^2 G}{c^2 r^3} = \frac{c\hbar}{r} \qquad\qquad r = \sqrt{\frac{\hbar G}{c^3}}$$

after making sure that both sides of the equation have the same dimensions. This is the Planck length. Thus, the electromagnetic, the weak force, the strong force and the gravitational force all become comparable somewhere near the Planck scale. The question is now why is $f$ so small. All the other not yet discovered particles could have mass in the order of the Planck scale.

# 3 Supersymmetry and Grand Unification

## 3.1 Renormalization of the Mass of the Higgs Boson

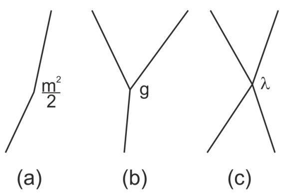
Renormalization is a combination of two things. One part is learning how to eliminate out of the description of physics distances which are so small that they are irrelevant to the questions one is asking. The other part is dimensional analysis which tells one the answer to very difficult problems in Quantum Field Theory. The nucleus, for example, consists of quarks but often can simply be seen as a collection of protons and neutrons. Quantum Chromodynamics is needed to find the properties of protons and neutrons such as mass and spin as well as the forces between them, but then it is no longer interesting what they are made out of. They also move with low speed such that there is no need for relativity theory, and, similarly, if one is interested in atoms, it may not be important to know what makes up a nucleus. If one knows the different nuclei with their mass and charge for the different elements and isotopes plus the electrons one can just study the various atoms. If one only cares about the physics of molecules one can ignore the structure of atoms to a certain degree. Ignoring fine-grained details gives a coarse-grained description which may not be exact but is more useful for the purposes.

The same is true in Quantum Field Theory. There are things at all possible scales, there are waves with arbitrarily small wave length and so on. Thus, one invents a way to summing up all these things which are too small to be interesting for a certain question and replaces them with new effective parameters. This is all renormalization really provides, together with dimensional analysis.

Starting from nuclei and electrons to atoms, for example, and ignoring small and therefore fast degrees of freedom, the dynamics of two atoms can be studied. The electrons are very fast in comparison to the nuclei. Because the lightest nucleus is two-thousand times heavier than the electron, and one can look at the atom with the nucleus as a very slowly moving bowling ball and with the electrons as fast flying flies. In a first approximation, one can assume that the nucleus does not move at all. One starts with a Hamiltonian as usual in Quantum Mechanics. The problem is not relativistic, and the two atoms are assumed to be the same. Thus, the kinetic energy of the two nuclei is $p_1^2/2M$ and $p_2^2/2M$, and their potential energy is the Coulomb potential $e^2/R_{12}$. The terms coming from the electrons are their kinetic energy $\sum q_i^2/2m$ plus the Coulomb potential $e^2/r_{ij}$ between them. The Hamiltonian also contains Coulomb terms for the attraction between the electrons and the two nuclei. Using $E_{\text{electrons}}$ for all the terms in the Hamiltonia where the electrons are involved, this is a function of the positions of the two nuclei $E_{\text{electrons}}(R_{12})$ or their distance $R_{12}$ which can be assumed to be fixed. This results from solving the Schrödinger equation for the lowest energy eigenvalue of the electrons in the background of the two nuclei. Thus, the problem has now been reduced to a Hamiltonian only for the nuclei and their positions, and the potential energy has been renormalized to the effective potential energy $e^2/R_{12} + E_{\text{electrons}}(R_{12})$.

Three units are needed in physics, but one can get rid of two by setting $c = \hbar = 1$. One unit is still needed and can be a length, an energy, a time and so on. Mass, energy and momentum have the same units, and time and length have the same units, but mass and time have inverse units.

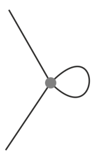
One can apply renormalization to a typical quantum field theory, for example, one with a single scalar field $\Phi$ which has a Lagrangian $\mathcal{L} = (\partial_\mu \Phi)^2 - V(\Phi)$ from which one can derive some Feynman diagrams. The action $S = \int d^4x\, \mathcal{L}$ has units of $\hbar$ and is therefore just a dimensionless number because $\hbar$ has been set to one. The dimension of $\Phi$ can be determined by dimensional analysis. The dimension of $(\partial_\mu \Phi)^2$ must be the inverse of length to the power of four to compensate the integral over $d^4x$. From this follows that $\Phi$ has the dimension of inverse length. This is an example of dimensional analysis.




The potential energy $V(\Phi)$ contains terms like $\frac{m^2}{2}\Phi^2$, $g\Phi^3$, $\lambda\Phi^4$ and so on where $\lambda$ is dimensionless. The Feynman diagrams (a), (b), (c) correspond to the terms with $\Phi^2$, $\Phi^3$, $\Phi^4$, respectively. The mathematical meaning of the propagator, which is shown in Feynman diagrams as straight lines from point $x$ to point $y$, is the amplitude $\langle 0|\Phi(y)\Phi(x)|0\rangle$ that the particle is detected at $y$ when it has been created at $x$. The question is what the dimension of this amplitude is. The vacuum has no dimension. A state is just a specification of a configuration and has no dimension, but $\Phi$ has dimension inverse of length. Thus, one can guess

$$\langle 0|\Phi(y)\Phi(x)|0\rangle = \frac{1}{|x-y|^2} \tag{3.1}$$

from dimensional analysis if there is no mass in the problem. If the particle has no mass, then the propagator is just one over the distance between the points squared. It blows up when the distance between the points is very small. That is the source of all divergences in Quantum Field Theory.



Renormalization of the mass which usually only appears squared in Quantum Field Theory has to do with the Feynman diagram in the above figure (a). A particle is absorbed and a particle is emitted. The point of absorption and the point of emission may be different as long as the distance between them is smaller than the scale of interest. Modeling this absorption and emission as a Feynman diagram of type (c) in the figure above, but where two propagators are replaced by one propagator as in the figure on the right side. For this propagator, the two points $x$ and $y$ are the same such that equation (3.1) results in a division by zero. If one is interested in length a little bit longer than the cutoff value $\delta$, one can assume that the two points $x$ and $y$ are separated by $\delta$ such that the propagator becomes $\lambda/\delta^2$. The mass consists now of the two terms $m_0^2/2$ which is the original mass term from the Feynman diagram (a) above and this additional value $\lambda/\delta^2$ where some numerical factors have been omitted. This means that the vertex in the Feynman diagram is smeared over some small region.

This renormalized mass is not complete because there are other terms from more complicated Feynman diagrams such as the one in the figure on the right side with the coupling constant proportional to $\lambda^2$ as it contains two vertices of type (c). If the lower vertex is assumed fixed one can sum or integrate over all the places where the upper vertex can be. If the lower vertex is called $x$ and the upper $x + \Delta$ then each of the three propagators between the two vertices is $1/\Delta^2$ because of (3.1) leading to $\int d^4\Delta\, \lambda^2/\Delta^6$. The integral goes from the cutoff value $\delta$ to infinity, and dimensional analysis shows that the integral can only depend on $\delta$ and must be $\lambda^2/\delta^2$ given that $\lambda$ is a dimensionless constant. Some numerical factors have also here been omitted. There are still more complicated Feynman diagrams with higher powers of $\lambda$ to be taken into account. Thus, there is an infinite series of terms each with $1/\delta^2$, but there are also terms of type (b) resulting from $\Phi^3$ and adding logarithmic terms multiplied by $g^2$ to the mass. The terms with $\delta^2$ in the denominator dominate and lead to an infinite sum of terms $m_0^2/2 + \lambda/\delta^2 + \lambda^2/\delta^2 + \lambda^3/\delta^2 + ...$ where the different terms have additional factors not shown here. This process is called mass renormalization and results in the effective mass.

However, not only the mass gets renormalized but any parameter in the Lagrangian such as $\lambda$, for example, can be renormalized. From an experimental point of view the Feynman diagram (c) corresponds to many diagrams where the vertex in the center is an area smaller than the cutoff value $\delta$. The result is renormalization of everything and the parameters of the theory that one measures are not the parameters that are the input to the theory.

This is renormalization of a scalar particle such as the Higgs boson. If the cutoff value $\delta$ is the Planck length, one is interested in a mass of about 200 GeV, but $\delta$ is in the order of $10^{-19}\,\mathrm{GeV}^{-1}$ and $1/\delta^2$ is in the order of $10^{38}$ GeV. Thus, the mass can only become the expected 200 GeV if the different contributions to the mass term which build a sum with an infinite number of positive and negative terms partially with huge values cancel nicely. This is called fine-tuning, and the corresponding puzzle is called the fine-tuning problem of the Higgs boson.

The Mexican hat potential of the Higgs boson has the form $-\frac{\mu^2}{2}\Phi^2 + \lambda\Phi^4$ plus possibly some higher-order terms which are believed not to be important. The minimum is $f^2 = \mu^2/\lambda$, and $f$ is in the order of 200 GeV. Thus, also $\mu$ must be in this order if $\lambda$ is in the order of one, but it is composed out of a ridiculous sum of terms every one of which may be 38 orders of magnitude bigger. This fine-tuning problem is also called gauge hierarchy problem and only exists for the Higgs boson.

## 3.2 Renormalization of the Mass of Other Particles

Thus, fermions do not have this problem. The mass term $m\bar{\Psi}\Psi$ in the Lagrangian similarly to the term in the Lagrangian of the Higgs boson also absorbs and emits a particle, but it either absorbs a left-handed particle and emits a right-handed particle, or absorbs a right-handed particle and emits a left-handed particle. If an electron, for example, emits a gauge boson, it does not change the handedness, but emitting a scalar particle does change the handedness. The process of emission and reabsorption of a photon within an area smaller than the cutoff value does not change the helicity, but also the emission and reabsorption of a scalar particle does not change the helicity outside of the cutoff region used for renormalization because it changes it an even number of times inside of the cutoff region. Thus, a massless electron or other fermion stays massless and does not get mass by mass renormalization.

A mass term means changing between left-handedness and right-handedness. Other than the scalar particle Higgs-boson, fermions cannot emit and absorb a scalar particle because it would change handedness an even number of times. There is therefore only one fine-tuning problem in the Standard Model, and this is renormalization of the mass of the Higgs-boson. An electron with starting mass $m_0$ may emit a photon, change handedness due to the mass term and reabsorbs the photon. This would add $e^2 m_0$, and the complete sum would be $m_0 + e^2 m_0 + e^4 m_0 + ...$ with also the more complicated Feynman diagrams. One might still ask why the electron is so light, but there is no fine-tuning problem because a small mass stays small. Gauge bosons have the same property. If they start massless they stay massless. Thus, only the Higgs boson has this fine-tuning problem. If $f$ is fixed then all the masses of the other particles are fixed. Supersymmetry and other theories are supposed to solve this fine-tuning problem of the Higgs boson, and there are theories where the mass of the Higgs boson is not driven to enormously large values by renormalization.

There is another extreme fine-tuning problem in physics related to gravity. It is the problem of renormalization of the vacuum energy. For most purposes in physics it does not matter what the energy of the vacuum is because it is an additive constant and only differences of energy are important. The right way to think about the energy of an electron, for example, is the extra energy one has to add to the vacuum when adding an electron whatever the mass of the vacuum is. The vacuum energy is also renormalized by Feynman diagrams with no input and no output particles which contribute typically $\lambda/\delta^4$. The vacuum energy does not matter until gravity comes into the picture because the source of the gravitational field is energy, and the vacuum gravitates if it has energy. This has to do with dark energy and the cosmological constant, and it also needs fine-tuning to a very high precision.

These are the two fine-tuning problems known in physics, and they are big puzzles in Particle Physics. It may turn out that they origin from a misstatement of the problem but it is a long time since these problems have been identified and nobody knows what the solution to them is. Supersymmetry is a potential solution to the fine-tuning of the Higgs boson but not to the fine-tuning of the cosmological constant.

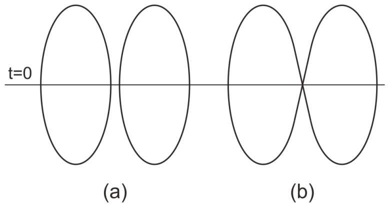
## 3.3 Some Differences Between Fermions and Bosons

It is generally assumed that rotation by $2\pi$ is equivalent to no rotation at all but this is not true. There is a topological theorem about rotation in space. If one takes a ball and connects it to the walls of a box by various strings then one gets a tangle by rotating the ball by $2\pi$. It is not possible to untangle it by moving around the strings which are assumed to be flexible, stretchable and unbreakable but without rotating the ball further. However, after a rotation by $4\pi$ (or generally by an even number of $2\pi$) this is always possible.

A similar situation exists in Quantum Mechanics. A wave function of a particle is assumed with fixed position because only the spin is of interest. It is described by a quantum state $|s\rangle$ which is half integer for fermions and integer for bosons. The question is what happens to the spin when the particle is rotated by $2\pi$. (The spin of an electron, for example, can be rotated using a magnetic field.) One can assume that rotating it by $2\pi$ multiplies the state by a number $|s\rangle \to \xi\,|s\rangle$. (The reason for not using an operator or so is that physicists secretly believe that no experiment will detect a rotation by $2\pi$.) Rotating again by $2\pi$ gives $\xi^2\,|s\rangle$. Since rotating by $4\pi$ is equivalent to doing nothing at all one can conclude $\xi^2 = 1$ and therefore $\xi = +1$ or $\xi = -1$.

One might take a big number of electrons all prepared the same way and then rotate one half by $2\pi$ and leave the other half untouched. Afterwards it would not be possible to detect a difference. However, with a two-slit experiment the change of the sign can be detected. Placing a magnetic field behind each slit will cause the electron to be trapped for a while. If one rotates one of the two magnetic field but not the other while the electron is trapped then the wave function may or may not get multiplied by a minus sign. If it gets multiplied by $-1$ the interference pattern gets shifted by half a wavelength. The result is that all fermions get a minus sign, and all bosons get a plus sign when rotated by $2\pi$. (This experiment can actually be performed.)

In the following simple Feynman diagrams in the form of loops are considered which are part of the vacuum structure. A loop is the production of a particle-antiparticle pair and their annihilation where the particles can be fermions or bosons, or alternatively it can be thought of as a particle going around in a closed loop in spacetime. Each diagram has a value which is a contribution to the vacuum energy and the question is whether it shifts it up or down.




The sign associated with one loop can either be positive or negative, and two loops as in the figure (a) are multiplied because two Feynman diagrams contribute just their product. Thus, the value associated with the two loops is always positive independent of whether a single loop is positive or negative. If the two loops are fermions, and if the two loops are close together, the two fermions may change their place as in (b) and the amplitude changes sign because the interchange of two fermions give a minus sign while the interchange of two bosons gives a plus sign. (It is assumed that at time $t = 0$ both pairs have been created but not yet annihilated.) If the two loops are bosons, both situations in the figure give a positive contribution, but if the two loops are fermions, the situation (a) gives a positive and the situation (b) which is actually only one loop gives a negative contribution. Therefore, the contribution of a loop of fermions to the vacuum energy is negative, and the contribution of a loop of bosons is positive.

The vacuum energy which is called dark energy is extremely small, but the naive direct calculations give a result an order of magnitude of 123 larger. There is a lot of cancellations between bosons and fermions, but there must be a boson for each fermion, both with the same mass and other properties, to cancel the vacuum energy out completely, because the Feynman diagrams for these loops depend on the mass of the particles. The fact that they cancel out to 123 decimal places is bizarre, but at least there is a possibility for positive and negative contributions canceling out.

## 3.4 Short Introduction to Supersymmetry

If there would be a theory which is sufficiently symmetric between bosons and fermions this would give a chance for canceling out all this vacuum energy. Supersymmetry is basically the idea that for every boson there is a fermion with the same mass and charge and vice versa such that the contributions to the vacuum energy cancels out. The mass of the particles with opposite statistics cannot be exactly the same as one knows from experimental facts, but the cancellation can be strong enough. The superpartners of the electron and positron, for example, have never been detected and this means, if they really exist, they must be very heavy. If an electron and a positron collide and build a virtual photon, this photon can split into the boson pair corresponding to the electron and positron as their partners. The fact that this has never been detected and that not even one partner of the known particles has been detected in the laboratory might just mean that no experiment so far had enough energy.

As shown above the renormalized mass of a Higgs boson requires that the self-energy terms of the various Feynman diagrams cancel to a high precision while the renormalization of the mass of fermions does not have this problem. Thus, if there is a theory such that even after all the renormalization and the worst looking Feynman diagrams there is enough symmetry between bosons and fermions which makes certain that the bosons and fermions travel in pairs and that the partners have exactly the same mass then the bosons do not get any mass if they started without because also the fermions do not get any mass through renormalization if they started without mass.

Thus, if there is a fermion for every boson and vice versa as the partner with the same electric charge and with the same mass then this would double the spectrum of all the elementary particles. It is known that the masses cannot be exactly the same and in fact they must be quite different. However, there may be enough symmetry that the different self-energy terms cancel good enough that one does not get this explosively large answer. Supersymmetry is among other things the idea that every particle has a partner of the opposite statistics. Particles come therefore in so-called superpartner pairs. If superpartners have the exact same mass the contributions to the vacuum energy would exactly cancel. With approximately the same mass the cancellation may be nearly exact. However, as mentioned above none of the superpartners have been detected in the laboratory yet.

The particles known today get their mass from the Higgs phenomenon, but other particles may get it from other sources. Thus, the fact that none of the superpartners have been detected yet as needed by these supersymmetric theories is not that weird anymore. Collisions with high enough energy may show some of these superpartners in the future. The superpartners, however, cannot be too heavy because the cancellation would otherwise be too poor and supersymmetry would not provide protection against these enormously large self-energy corrections. They cannot be heavier than a couple of hundred GeV, and this is within the experimental range of the Large Hadron Collider at CERN. There is a notion of spontaneously broken supersymmetry which creates mass differences for particles. If these superparticles exist, they are expected to be very instable and they would decay into know particles. There may be one of the superpartners that is very likely electrically neutral and that is stable. The superpartners have to decay to something and in many theories they have to decay to something that is still a superpartner. It may well be that this so-called lightest superpartner is dark matter and because it is electrically neutral it is not easy to detect it.

Superpartners of bosons typically have spin $\frac{1}{2}$ and their names end with ino such that the gravitino would be the superpartner of the graviton. Because the superpartners of fermions are usually scalar particles with spin 0 the letter s is attached in front such that the bosonic superpartner of the quark is a squark. Superpartners have always spins that differ by a unit of $\frac{1}{2}$. The existing ordinary particles such as the electron keep their names, of course, in the context of supersymmetry.

Supersymmetry may at most be a good approximation. If supersymmetry would be the correct description of physics and superpartners would have the exact same mass, the world would be completely different. The real world is certainly not supersymmetric in this sense. Chemistry as known would not work because in an atom with three or more electrons, one electron can emit a photino as the superpartner of the photon which would also be massless and turn into a selectron as the superpartner of an electron. Because it is a boson, it could stay in the lowest orbit because the Pauli exclusion principle does not prevent this. In a big atom all the outer electrons would eventually emit photinos to get rid of the extra energy of the outer orbits and turn into selectrons. The result would be a Bose condensate in the lowest energy level. Chemistry needs valence electrons, but there would be none.

## 3.5 Motivation for Supersymmetry

A lot of Quantum Field Theory and Particle Physics is dependent on calculating Feynman diagrams, and many Feynman diagrams govern scattering processes. One has to sum over all Feynman diagrams that go into the calculation. Another thing that Feynman diagrams do is to help calculating the effective Lagrangian which is an approximation for a Lagrangian that is too complicated to calculate exactly. An effective Lagrangian typically needs a cutoff for smallest distance or largest momentum. Propagators in Feynman diagrams are the amplitude for putting in a particle in one place and taking it out in another place. They are functions of relative coordinates $\Delta^\mu$ which is a 4-vector. They can also be Fourier transformed and expressed as a function of momentum. The Feynman rules are usually expressed in terms of momentum, and Feynman diagrams become integrals over momenta. The divergences can be thought of either as infinities which occur because of small distances or as infinities due to very large momenta. (Note that momentum is related to inverse wavelength.)

With $c = \hbar = 1$ as usual the units of mass is inverse length. Thus, if one specifies a mass one also specifies a length scale. When talking about massless particles, however, there is no length scale involved in the propagators other than the distance between $x$ and $y$. The amplitude for massless particles can be guessed by dimensional analysis because the action is always dimensionless. For a massless scalar field the action with only the kinetic term in the Lagrangian is $\int d^4x\, (\partial_\mu \Phi)^2$ such that one can conclude that the dimension of $\Phi$ is inverse length. Adding a mass term to the Lagrangian shows that the dimensions are still consistent. If one adds a third term $\lambda\Phi^4$ the coupling constant $\lambda$ must be dimensionless.

The propagator for a massless scalar particle with spin 0 is $\langle 0|\Phi(y)\Phi(x)|0\rangle$ where $\Phi$ is real if the particle is not electrically charged. With the distance $\Delta$ which is a 4-vector and $\Delta^2 = \Delta_\mu\Delta^\mu$ the propagator can only depend on $1/\Delta^2$ times a numerical factor. Because $E = \sqrt{p^2 + m^2}$ for particles with mass and $E = |p|$ for massless particles, the mass term becomes negligible for very large momenta corresponding to short wavelengths. Things like propagators become insensitive to the mass at very small distances. Thus, equation (3.1) is not only correct for massless particles with spin zero but also for particles with mass to very high accuracy. The real value would be $e^{-\Delta m}/\Delta^2$.

The action for fermions is $\int d^4x\, \overline{\Psi}\gamma^\mu\partial_\mu\Psi + m\overline{\Psi}\Psi$ and the dimensions of $\Psi$ are therefore inverse of length to the power of 3/2 because the action is dimensionless. The propagator for massless fermions $\langle 0|\Psi_j^\dagger(y)\Psi_k(x)|0\rangle$ where $\Psi^\dagger$ and $\Psi$ have four components due to the Dirac equation is therefore

$$\langle 0|\Psi_n^\dagger(y)\Psi_m(x)|0\rangle \sim \frac{\Delta_\mu\gamma^\mu_{nm}}{\Delta^4} \sim \frac{1}{\Delta^3} \tag{3.2}$$

without some numerical factors, and it is a bit smaller for fermions with mass.

Exchanging two bosons leaves the sign of the amplitude the same because of $\varphi(x_1, x_2) = \varphi(x_2, x_1)$ for the Schrödinger wave function while exchanging two fermions change the sign of the amplitude because of

$\psi(x_1, x_2) = -\psi(x_2, x_1)$ for the Schrödinger wave function. In other words, the amplitude for a one-loop Feynman diagram is positive for bosons and negative for fermions as discussed above. This helps to cancel the large values due to the infinite sum of diverging terms with $\delta^2$ in the denominator.

It has been shown in the discussion of renormalization that the Higgs field gets terms of the form $\lambda/\delta^2$ added to the mass term. The Feynman diagram where the Higgs boson splits into two fermions which later join again giving back the Higgs boson on the other hand contains the propagators of the two fermions leading to a factor of the form $1/\Delta^6$ due to (3.2) or to $-g^2/\delta^2$ with the dimensionless coupling constant $g$ inserted and after integration over spacetime from $\delta$ to infinity. Thus, both contributions together are $\lambda/\delta^2 - g^2/\delta^2$. It would be very surprising if nature had made $\lambda = g^2$ for these two independent coupling constants without some symmetry or other reason, and supersymmetry gives such a mathematical reason why this equation is fulfilled and is not just an accidental coincidence.

## 3.6 Some General Properties of Symmetries

The question whether supersymmetry is a continuous symmetry or not has no answer. A symmetry in general has symmetry operations which are applied to state vectors $|\varphi\rangle$. If $\mathbf{U}$ is a symmetry operator, then it must be unitary such that $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$ because this ensures conservation of probability. A continuous group can be built out of infinitesimally small changes. The generators $\mathbf{L}_j$ of the symmetry group of rotations with $\mathbf{I} + i\varepsilon\mathbf{L}_j$ are the angular momenta about the different axes, and the generators $\mathbf{p}_j$ of the symmetry group of translations with $\mathbf{I} + i\varepsilon\mathbf{p}_j$ are the momenta in the direction of the different axes. The commutator algebra of the generators for a continuous symmetry represents the whole group.

If $\mathbf{U}_1$ and $\mathbf{U}_2$ are two rotations about different axes then $\mathbf{U}_2^\dagger\mathbf{U}_1^\dagger\mathbf{U}_2\mathbf{U}_1 \neq \mathbf{I}$ but some other rotation. (If the two rotations are about the same axes then the result would be the identity.) For infinitesimal rotations about the $x$- and the $y$-axis, for example, this would give $(\mathbf{I} - i\varepsilon_2\mathbf{L}_y)(\mathbf{I} - i\varepsilon_1\mathbf{L}_x)(\mathbf{I} + i\varepsilon_2\mathbf{L}_y)(\mathbf{I} + i\varepsilon_1\mathbf{L}_x)$ in which the terms first order in $\varepsilon_1$ and $\varepsilon_2$ cancel. The important terms that remain are $\varepsilon_1\varepsilon_2(\mathbf{L}_x\mathbf{L}_y - \mathbf{L}_y\mathbf{L}_x)$ or $\varepsilon_1\varepsilon_2[\mathbf{L}_x, \mathbf{L}_y]$ together with $\mathbf{I}$. Thus, if the commutator $[\mathbf{L}_x, \mathbf{L}_y]$ is zero then $\mathbf{U}_2^\dagger\mathbf{U}_1^\dagger\mathbf{U}_2\mathbf{U}_1 = \mathbf{I}$ for the two infinitesimal rotations $\mathbf{U}_1 = \mathbf{I} + i\varepsilon_1\mathbf{L}_x$ and $\mathbf{U}_2 = \mathbf{I} + i\varepsilon_2\mathbf{L}_y$. This is the geometric meaning of the commutator algebra $[\mathbf{L}_j, \mathbf{L}_k] = i\,\varepsilon_{jkl}\,\mathbf{L}_l$ for the group of rotations, and $\mathbf{U}_2^\dagger\mathbf{U}_1^\dagger\mathbf{U}_2\mathbf{U}_1 = \mathbf{I} + i\varepsilon_1\varepsilon_2\mathbf{L}_z$ in the example. The commutator algebra for translations is $[\mathbf{p}_j, \mathbf{p}_k] = 0$ as one can easily see in Euclidean space, but the generators of rotations and of translations do not commute.

In general, the commutators keep track of the fact that group elements applied in different orders do not cancel out. The generators $\mathbf{G}_i$ of a continuous group build a Lie algebra $[\mathbf{G}_j, \mathbf{G}_k] = i\,c_{jkl}\mathbf{G}_l$ where all commutators are linear combinations of the generators and where the factors $c_{jkl}$ are called the structure constants of the group. Thus, the Lie algebra is closed. The generators are usually hermitian and have therefore real eigenvalues, but sometimes it makes sense to build complex combinations of generators such as the raising and lowering operators $\mathbf{L}_x \pm i\mathbf{L}_y$ for the angular momentum.

Because the energy state of a system stays the same when a symmetry operation $\mathbf{U}$ is applied, an eigenvector of $\mathbf{H}$ is also an eigenvector of $\mathbf{UH}$. Therefore, the commutator of $\mathbf{U}$ and $\mathbf{H}$ is zero. When every symmetry operator commutes with the Hamiltonian, also the generators commute with the Hamiltonian. The time derivative of an operator that commutes with the Hamiltonian is zero. Therefore, symmetries and conservation laws imply each other.

## 3.7 Grassmann Numbers

All the symmetries discussed so far give back bosons when acting on bosons and give back fermions when acting on fermions. Actually, all symmetries discussed so far do not change the spin, the charge and the mass. Supersymmetry is a new kind of symmetry that changes bosons to fermions and vice versa, and it changes therefore also the spin. The mathematics is very weird, and it is impossible to visualize it. A theory with this symmetry has a fermion for every boson and a boson for every fermion with exactly the same mass. The generators of supersymmetry are called $\mathbf{Q}$, and they turn a boson into a fermion $\mathbf{Q}\,|b\rangle = |f\rangle$ and a fermion into a boson $\mathbf{Q}\,|f\rangle = |b\rangle$. This means that $\mathbf{Q}$ annihilates the boson and creates a fermion or vice versa. With $\mathbf{Q} = a^+c^- + c^+a^-$ where $a^-$ and $a^+$ are the annihilation and creation operators for bosons and $c^-$ and $c^+$ are the annihilation and creation operators for fermions, one of the terms of $\mathbf{Q}$ does nothing and the other changes one kind into the other. When $\mathbf{Q}$ acts on the vacuum $|0\rangle$, it gives 0 because it tries to annihilate a boson and a fermion first, but there is none.

When $\mathbf{Q}$ acts on a particle with spin zero, it produces a fermion with spin $\frac{1}{2}$. But does it create a spin up or a spin down? Thus, there must be two operators $\mathbf{Q}_1$ and $\mathbf{Q}_2$ or, there must be multiple $\mathbf{Q}_i$ in general. The creation and annihilation operators for bosons are described in terms of commutators, and the creation and annihilation operators for fermions are described in terms of anticommutators.

Large groups of bosons behave classically, but one cannot build large groups of fermions. Bosons use the normal arithmetic operations where ordinary numbers commute, but fermions need an anticommuting generalization for numbers which are called Grassmann numbers. Ordinary numbers commute such that $\alpha_i\alpha_j = \alpha_j\alpha_i$, and Grassmann numbers anticommute such that $\theta_i\theta_j = -\theta_j\theta_i$. The square of a Grassmann number is therefore zero because $\theta^2 = \theta\theta = \frac{1}{2}(\theta\theta + \theta\theta) = 0$. Mixing ordinary numbers and Grassmann numbers follows the rules $[\alpha_i, \alpha_j] = 0$, $[\alpha_i, \theta_j] = 0$, and $\{\theta_i, \theta_j\} = 0$. An ordinary number multiplied with a Grassmann number gives a Grassmann number, but Grassmann number times Grassmann number gives an ordinary number. Ordinary numbers and Grassmann numbers can be real or complex. The Grassmann number $\bar{\theta}$ is the complex conjugate of the Grassmann number $\theta$. Ordinary numbers can be called even and Grassmann numbers odd.

Functions of Grassmann numbers can be written as power series. Assuming there is only one Grassmann variables $\theta$, then there are only linear polynomials of the form $F(\theta) = A+B\theta$. Now assuming there are two Grassmann variables $\theta_1$ and $\theta_2$, then the most complex function is $F(\theta_1, \theta_2) = A + B_1\theta_1 + B_2\theta_2 + C\theta_1\theta_2$. Thus, there are also for more variables only a limited number of polynomial functions. When writing a function with Grassmann variables one assumes that all terms are either odd or even. If, for example, $F(\theta_1, \theta_2)$ is an even function, then $A$ has to be even, the $B_i$ have to be odd such that $B_i\theta_i$ is even, and $C$ again must be even.

One can define differentiation and integration of functions of Grassmann numbers. It turns out that differentiation and integration are the same thing. The partial differential of a function is the same as for ordinary numbers with the exception that a derivative passing through a Grassmann number changes the sign. For a function with one variable $\partial/\partial\theta\ (A + \theta B) = B$, but $\partial/\partial\theta\ (A + B\theta) = \pm B$ depends on $B$. If $A$ is a Grassmann number then $B$ must be an ordinary number such that $B\theta = \theta B$. If on the other hand $A$ is an ordinary number then $B$ must be a Grassmann number such that $A + \theta B = A - B\theta$. The derivative $\partial/\partial\theta\ (A+\bar{B}\theta+\bar{\theta}B+C\bar{\theta}\theta)$ is $-\bar{B} - C\bar{\theta}$ for an even $A$ and $\bar{B}+C\bar{\theta}$ for an odd $A$. Differentiation of Grassmann functions also satisfies the product rule.

Finally, the integral is $\int F(\theta)\, d\theta = \frac{\partial}{\partial\theta}F(\theta)$ if the integration by parts formula is true since the rules for integration follow from linearity and from integration by parts, and the integral is always a definite integral. Because there are so few functions with Grassmann variables, all integrals can be listed in a table. The simplest integral is $\int d\theta = 0$ followed by $\int \theta\, d\theta = 1$ as a convention. Combined and using linearity the integral of the most general function with one Grassmann variable is $\int(A + B\theta)\, d\theta = B$ and the one with two Grassmann variables $\theta$ and $\bar{\theta}$ is $\int(A + \bar{\theta}\psi + \bar{\psi}\theta + B\bar{\theta}\theta)\, d\theta d\bar{\theta} = B$.

The commutator is $[\frac{\partial}{\partial x}, x] = \mathbf{I}$ because $[\frac{\partial}{\partial x}, x]F(x) = F(x) + xF'(x) - xF'(x) = F(x)$ for any function $F(x)$ with an ordinary variable. This is the commutator relation known from Quantum Mechanics for the position and momentum operators. For functions with Grassmann variables the equivalent relation is the anticommutator equation $\{\frac{\partial}{\partial\theta}, \theta\} = \mathbf{I}$ . To demonstrate this relation for one kind of function, the even function $F(\theta) = A + B\theta$ is used and gives $\{\frac{\partial}{\partial\theta}, \theta\}F(\theta) = B\theta + A = F(\theta)$.

## 3.8 Simplified Model of Supersymmetry

Supersymmetry is a very bizarre symmetry of spacetime, and one has to extend spacetime to something radically new. Fermion fields are odd, but one never measures fermion fields because measurable quantities are always ordinary real numbers. The things one measures in nature are things one should think of as even elements if there is a reason at all to think about a Grassmann algebra. In ordinary symmetries, the generators build a commutator algebra. In supersymmetries, the generators build an anticommutator algebra and are odd elements.

The two generators of supersymmetry are $\mathbf{Q}$ and the complex conjugate $\bar{\mathbf{Q}}$. They are odd elements such that $\{\mathbf{Q}, \mathbf{Q}\} = \{\bar{\mathbf{Q}}, \bar{\mathbf{Q}}\} = 0$. The most primitive version of a supersymmetry introduced here is related to the full supersymmetry the same way as the harmonic oscillator is related to Quantum Field Theory. The bosonic even creation and annihilation operators $a^\pm$ satisfies the algebra $[a^+, a^+] = [a^-, a^-] = 0$ and $[a^-, a^+] = 1$, and the fermionic odd creation and annihilation operators $c^\pm$ satisfy the algebra $\{c^+, c^+\} = \{c^-, c^-\} = 0$ and $\{c^+, c^-\} = 1$. In addition these creation and annihilation operators satisfy the mixed commutator relations $[a^\pm, c^\pm] = 0$. These operators allow to talk about bosons and fermions at rest which can either be there or not be there but without considering their motion. They are assumed to have all the same mass $m$ such that the total energy is $m$ times the number of particles.

The generators of the supergroup are defined as $\mathbf{Q} = \sqrt{m}\, a^+ c^-$ which removes a fermion and creates a boson as well as $\bar{\mathbf{Q}} = \sqrt{m}\, a^- c^+$ which adds a fermion and removes a boson. It does not matter whether the particle of one kind is first removed and the particle of the other kind created afterwards or the other way around. However, to replace two fermions by two bosons, one cannot use $\mathbf{QQ}$ because $\mathbf{Q}^2 = 0$ due to the Grassmannian characteristic.

With $\{\mathbf{Q}, \bar{\mathbf{Q}}\} = m(a^+c^-\, a^-c^+ + a^-c^+\, a^+c^-) = m(a^+a^-\, c^-c^+ + a^-a^+\, c^+c^- + a^+a^-\, c^+c^- - a^+a^-\, c^+c^-)$ where a term has been added and subtracted the anticommutator becomes $\{\mathbf{Q}, \bar{\mathbf{Q}}\} = m(a^+a^- + c^+c^-)$ because $a^+a^-(c^-c^+ + c^+c^-) = a^+a^-\{c^+, c^-\} = a^+a^-$ and $(a^-a^+ - a^+a^-)c^+c^- = [a^-, a^+]c^+c^- = c^+c^-$. Since $a^+a^-$ is the number of bosons $n_\mathrm{B}$ and $c^+c^-$ is the number of fermions $n_\mathrm{F}$ the anticommutator $\{\mathbf{Q}, \bar{\mathbf{Q}}\} = m(n_\mathrm{B} + n_\mathrm{F}) = \mathbf{H}$ is the total energy and therefore the Hamiltonian $\mathbf{H}$ which is the generator of time invariance. With $\{\mathbf{Q}, \mathbf{Q}\} = \{\bar{\mathbf{Q}}, \bar{\mathbf{Q}}\} = 0$ from above and, if the mass of bosons and fermions are equal as assumed here, the commutator $[\mathbf{Q}, \mathbf{H}] = 0$, because removing a particle at rest and adding another particle also at rest with the same mass does not change the energy, the algebra is complete. This means that $\mathbf{Q}$ is a conserved quantity, and this is surprising because $\mathbf{Q}$ is an odd quantity.

There is an even more exotic way to think about this supersymmetry because there are new dimensions of spacetime. In this simplified supersymmetry there is no space but only time because there is energy and the relation between time and energy is $i\frac{\partial}{\partial t}|\psi\rangle = \mathbf{H}|\psi\rangle$ or, equivalently, $e^{-i\mathbf{H}\delta}|\psi(t)\rangle = |\psi(t+\delta)\rangle$. The symmetry here is mixing up spacetime with identity of particles and operators changing particles into other particles, because the anticommutator of two of these supersymmetry generators amounts to a short translation in time. The symmetry of color in Quantum Chromodynamics or the symmetry of isospin also mix up particles, for example, by turning an up quark into a down quark, but they never shift time or space or have in general anything to do with spacetime. Here are two operators which give a small shift in time when they anticommute. (Note that spacetime is here simply time.) There is a way to think about this in terms of coordinate shifts, but it is not ordinary coordinates but Grassmann coordinates. Given $\psi(\theta, \bar{\theta}, t)$ as the wave function of some kind of strange superparticle located at Grassmann coordinates $\theta$ and $\bar{\theta}$. One symmetry is a shift of time from $t$ to $t+\delta$ which changes $\psi(\theta, \bar{\theta}, t)$ to $\psi(\theta, \bar{\theta}, t) + \delta\frac{\partial}{\partial t}\psi(\theta, \bar{\theta}, t)$. An analog of a shift in space allows one to invent the symmetry $\theta \to \theta + \varepsilon$ and $t \to t + i\bar{\theta}\varepsilon$. The effect on $\psi$ is $\psi \to \psi + \varepsilon[\frac{\partial}{\partial\theta} - i\bar{\theta}\frac{\partial}{\partial t}]\psi = \psi + \varepsilon\mathbf{Q}\psi$ where $\mathbf{Q} = \frac{\partial}{\partial\theta} - i\bar{\theta}\frac{\partial}{\partial t}$ is just a definition so far. With the anticommutator $\{\frac{\partial}{\partial\theta} - i\bar{\theta}\frac{\partial}{\partial t}, \frac{\partial}{\partial\bar{\theta}} + i\theta\frac{\partial}{\partial t}\} = -i\frac{\partial}{\partial t}$ which is $\mathbf{H}$ (possibly with some missing numerical factors in the anticommutator) this is a representation of $\mathbf{Q} = \frac{\partial}{\partial\theta} - i\bar{\theta}\frac{\partial}{\partial t}$ and $\bar{\mathbf{Q}} = \frac{\partial}{\partial\bar{\theta}} + i\theta\frac{\partial}{\partial t}$. However, this representation is not in terms of creation and annihilation operators but as operations in a fictitious new kind of spacetime that involves anticommuting coordinates. There is a geometric structure involving spaces which contain not only ordinary coordinates but also Grassmann coordinates. Moving around in this space entails moving around in ordinary spacetime with the generators momentum and energy and also transformations which move around in this Grassmann space with the supersymmetry generators $\mathbf{Q}$ and $\bar{\mathbf{Q}}$. These symmetries are called the supersymmetry, and the generalized space with Grassmannian dimensions is called superspace.

### 3.9 More on Supersymmetry

Supersymmetry is first of all a symmetry which takes fermions to bosons. It can be represented rather simply by operations which replace fermions by bosons. The unusual fact here is that the operations close only if one includes the Hamiltonian $\mathbf{H}$ showing that the symmetry operations not only mix fermions and bosons but when one anticommutes them one gets a translation in time. Thus, the supersymmetry generators are translations in the Grassmann space together with some translations in time.

The generators of angular momentum are infinitesimal rotations about an axis, and applying the same generator multiple times allows rotations about this axis for larger angles. The supersymmetry generator

$\mathbf{Q}$ is different because it allows to remove one boson and replace it with a fermion (or vice versa), but it is not possible to remove multiple bosons and replace them with fermions using $\mathbf{Q}^n$ because $\mathbf{Q}^2 = 0$.

There are several reasons for introducing supersymmetry. Firstly, supersymmetry controls the divergent self-energy (or self-mass) of the Higgs particle. Secondly, it gives candidates for new and not yet detected particles which are called superpartners and cannot disappear because a superpartner always decays into an ordinary particle and a superpartner. (As mentioned above the lightest superpartner which is supposed to be stable may be responsible for dark matter.) Thirdly, the extrapolation of the inverse of the running coupling constants for the emission of a photon, a W-boson and a gluon cross more exactly in one point at very high energy around $10^{16}$ GeV when the new Feynman diagrams involving superpartners are also taken into account than with only the known particles in the Standard Model of Particle Physics. The split into three different coupling constants is assumed to be the result of spontaneous symmetry breaking. (Also the coupling constant for gravity crosses at a point fairly close to this point.)

Spacetime in supersymmetry has the ordinary four dimensions $x^\mu$ and two complex (or four real) Grassmann space dimensions $\theta_\alpha$ which are very small because their square is zero. Superfields are functions of all these dimensions $\Phi(x, \theta)$. When expanded as $\Phi(x, \theta) = \varphi(x) + \theta_\alpha \psi_\alpha(x) + \theta_\alpha \theta_\beta Z_{\alpha\beta}(x) + \theta_1\theta_2\theta_3\theta_4 D(x)$ it shows that the field consist of several fields $\varphi(x)$, $\psi(x)$, $Z(x)$, $D(x)$ where some are boson fields and some are fermion fields depending one the oddness or evenness. The Lagrangian $\mathcal{L}(x, \theta)$ integrated into $\int \mathcal{L}(x, \theta)\, d^4x\, d^4\theta$ gives the action as usual. The question is just what the integral over Grassmann variables means. Only definite integrals are defined which are integrals over the entire, but due to $\theta^2 = 0$ very small $\theta$-space. The justification of the integration of Grassmann variables is their usefulness.

### 3.10 Construction of Supersymmetric Theories

With $\varphi$ for bosons and $\psi$ for fermions, the Lagrangian for bosons is $\mathcal{L} = \frac{1}{2}(\dot{\varphi}^2 - m^2\varphi^2)$ with the solution of the Euler-Lagrange equation $\ddot{\varphi} = -m^2\varphi$ that oscillates with frequency $m$. The Lagrangian for fermions is in some sense simpler, but it uses Grassmann numbers. Assuming a Lagrangian $\mathcal{L} = \dot{\varphi}\varphi$ for a boson, the result of $\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\varphi}} = \frac{\partial\mathcal{L}}{\partial\varphi}$ is just $\dot{\varphi} = \dot{\varphi}$. For a fermion, the solution for $\mathcal{L} = \dot{\psi}\psi$ is $\dot{\psi} = -\dot{\psi}$ or $\dot{\psi} = 0$. In order to make the fermions also oscillate, one can change the Lagrangian to $\mathcal{L} = \dot{\psi}\psi + m\psi^2$, but that does not work because $\psi^2 = 0$. Another try is $\mathcal{L} = \frac{1}{2}i(\dot{\psi}_1\psi_2 + \dot{\psi}_2\psi_1) + m\psi_1\psi_2$ with the solutions $i\dot{\psi}_1 = m\psi_1$ (or $i\dot{\psi}_1 = -m\psi_1$?) and $i\dot{\psi}_2 = m\psi_2$ which are oscillations. This gives the Dirac equation

$$\mathcal{L} = \frac{1}{2}\left(\dot{\varphi}^2 - m^2\varphi^2\right) + \frac{i}{2}\left(\dot{\psi}^*\psi + \dot{\psi}\psi^*\right) + m\psi^*\psi \tag{3.3}$$

for the Lagrangian with only time but no space direction and with $\psi = \psi_1 + i\psi_2$ an $\psi^* = \psi_1 - i\psi_2$.

Defining the supergenerators for supersymmetry a bit differently than above one gets $\bar{\mathbf{Q}} = a^+c^-\sqrt{2m}$ and $\mathbf{Q} = a^-c^+\sqrt{2m}$ with $\{\bar{\mathbf{Q}}, \mathbf{Q}\} = 2m(a^+a^- + c^+c^-) = 2\,\mathbf{H}$ in the conventional notation. The commutators $[\mathbf{Q}, \mathbf{H}] = [\bar{\mathbf{Q}}, \mathbf{H}] = 0$ show that $\mathbf{Q}$ and $\bar{\mathbf{Q}}$ are conserved quantities and are therefore called supercharges. The three quantities $\mathbf{Q}$, $\bar{\mathbf{Q}}$ and $\mathbf{H}$ build a closed algebra.

In Quantum Mechanics $\mathbf{H} = i\frac{\partial}{\partial t}$ and $\mathbf{p} = -i\frac{\partial}{\partial x}$ are related to derivatives with respect to time and space, respectively, and other symmetry generators (preserved quantities) are related to some combinations of derivatives as well. Also $\mathbf{Q}$ and $\bar{\mathbf{Q}}$ are related to the derivatives with respect to $\theta$ and can be written as

$$\bar{\mathbf{Q}} = \frac{\partial}{\partial\bar{\theta}} + i\theta\frac{\partial}{\partial t} \qquad\qquad \mathbf{Q} = \frac{\partial}{\partial\theta} + i\bar{\theta}\frac{\partial}{\partial t} \tag{3.4}$$

where part is convention, but the two plus signs must be the same such that $\{\bar{\mathbf{Q}}, \mathbf{Q}\} = 2i\frac{\partial}{\partial t}$.

For simplicity the following thoughts assume as above that there are no spatial dimensions, that $\varphi(t)$ is bosonic, and that $\psi(t)$ is fermionic. Thus, a superfield $\Phi$ can be constructed as

$$\Phi(t, \theta, \bar{\theta}) = \varphi(t) + \bar{\theta}\psi(t) + \bar{\psi}(t)\theta + D(t)\bar{\theta}\theta$$

expanded into a power series. The transformations $\theta \to \theta + \xi$, $\bar{\theta} \to \bar{\theta} + \bar{\xi}$, $t \to t + i\bar{\xi}\theta + i\xi\bar{\theta}$ preserve this simplified supersymmetry, and they lead to the change

$$\delta\Phi = \xi\frac{\partial\Phi}{\partial\theta} + \bar{\xi}\frac{\partial\Phi}{\partial\bar{\theta}} + \frac{\partial\Phi}{\partial t}(i\xi\bar{\theta} + i\bar{\xi}\theta) = \xi\left[\frac{\partial}{\partial\theta} + i\bar{\theta}\frac{d}{dt}\right]\Phi + \bar{\xi}\left[\frac{\partial}{\partial\bar{\theta}} + i\theta\frac{d}{dt}\right]\Phi = \left[\xi\mathbf{Q} + \bar{\xi}\bar{\mathbf{Q}}\right]\Phi$$

according to these shifts. This is like momenta but for shifts in the Grassmann coordinates together with an additional shift in time. This is the basic symmetry called supersymmetry, and it can be thought of as a geometric symmetry of shifting and translating in this space with Grassmannian dimensions in addition to time.

The action is $\int dt\, d\theta\, d\bar{\theta}\, \Lambda$ for the super-Lagrangian $\Lambda$ and using the Grassmannian coordinates similarly to ordinary spatial coordinates. The super-Lagrangian $\Lambda(t,\theta,\bar{\theta})$ is itself a superfield like $\Phi(t,\theta,\bar{\theta})$ and has a similar expansion into a power series. The integral over the Grassmann variables picks out the last term which is also called the D-term. Calling this term $\Lambda_{\bar{\theta}\theta}$, the ordinary Lagrangian becomes $\int dt\, \Lambda_{\bar{\theta}\theta}$. This is the Lagrangian of a free field without interactions between bosons and fermions.

To create other superfields from a superfield, one can, for example, square it. Thus, one gets

$$\Phi^2 = \Phi \times \Phi = \varphi^2 + 2\bar{\theta}\varphi\psi + 2\varphi\bar{\psi}\theta + 2\varphi D\bar{\theta}\theta - 2\bar{\psi}\psi\bar{\theta}\theta$$

which might be a term to put into a super-Lagrangian and which would result as $2\varphi D - 2\bar{\psi}\psi$ in the Lagrangian after integration over the two Grassmann variables. This is not yet a very interesting term because it has no time derivatives. To get more interesting terms such as kinetic terms one cannot just use the time derivative because the result is no longer a superfield. Super-Lagrangians become more interesting and more realistic, when spacetime has the three spatial coordinates as well.

## 3.11 Supersymmetry in Full Spacetime

In the simplified supersymmetry above with only a time but no space dimensions there were relationships between the supergenerators $\mathbf{Q}$ and $\bar{\mathbf{Q}}$ and the Hamiltonian $\mathbf{H}$ but the energy belongs together with the momentum into a single 4-vector. The Dirac equation for massless fermions has two uncoupled components for left-handed and right-handed particles and the Hamiltonian is $\mathbf{H} = \sigma \cdot \mathbf{p}$ where left-handed particles have positive energy. This equation converted into a wave equation gives $i\partial_t \Psi = -i\sigma \cdot \partial_x \Psi$ or

$$i\sigma^\mu \frac{\partial \Psi}{\partial x^\mu} = 0$$

when adding the unit matrix as a fourth Pauli-matrix. This is the chiral Dirac-equation for a massless particle, and it is Lorentz-invariant.

This suggests that also the Grassmann parameters $\theta$ consist of two components similar to $\Psi$ such that the two entities $\theta_\alpha$ and $\bar{\theta}_\alpha$ are two-dimensional vectors with $\alpha \in \{1,2\}$. The expansion into a power series now goes up to the fourth power as the product of four distinct variables. The anticommutators and commutators $\{\bar{\mathbf{Q}}_\alpha, \mathbf{Q}_\beta\} = 2\,\sigma^\mu_{\alpha\beta}\mathbf{p}_\mu$, $\{\bar{\mathbf{Q}}_\alpha, \bar{\mathbf{Q}}_\beta\} = \{\mathbf{Q}_\alpha, \mathbf{Q}_\beta\} = 0 = [\mathbf{Q}_\alpha, \mathbf{p}_\mu] = [\bar{\mathbf{Q}}_\alpha, \mathbf{p}_\mu] = [\mathbf{p}_\mu, \mathbf{p}_\nu]$ build the superalgebra for the supersymmetry. As a generalization of (3.4)

$$\mathbf{Q}_\alpha = \frac{\partial}{\partial \theta_\alpha} + i\bar{\theta}_\beta \sigma^\mu_{\beta\alpha} \frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial \theta} + i\bar{\theta}\sigma\partial \qquad \bar{\mathbf{Q}}_\beta = \frac{\partial}{\partial \bar{\theta}_\beta} + i\sigma^\mu_{\beta\alpha}\theta_\alpha \frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial \bar{\theta}} + i\sigma\theta\partial \tag{3.5}$$

the operators $\mathbf{Q}_\alpha$ and $\bar{\mathbf{Q}}_\beta$ are a kind of momenta in the direction of the Grassmann coordinates. They operate on a superfield $\Phi(x,\theta,\bar{\theta})$ and give the small change in the superfield under a supersymmetric transformation. In other words, the variation $\delta\Phi$ of the superfield means either $\delta\Phi = \mathbf{Q}\Phi$ or $\delta\Phi = \bar{\mathbf{Q}}\Phi$.

A functions $F(x,y)$ is assumed with the constraint $\mathbf{D}F = 0$ meaning that the derivative with respect to $y$ is zero. An infinitesimal rotation $\delta F = i\mathbf{L}F$ where $\mathbf{L}$ is a generator of rotation changes $x$ components into $y$ components and vice versa. Thus, $\mathbf{D}\delta \neq 0$ and the rotated $F$ does not satisfy the same constraint as $F$. With $\mathbf{D}(F + \delta F) = \mathbf{D}(F + i\mathbf{L}F) = \mathbf{D}F + i\mathbf{D}\mathbf{L}F$ the term $\mathbf{D}F$ is zero and the term $\mathbf{D}\mathbf{L}F$ would be zero if $\mathbf{D}\mathbf{L} = \mathbf{L}\mathbf{D}$ because of $\mathbf{D}\mathbf{L}F = \mathbf{L}(\mathbf{D}F)$. A constraint like $\mathbf{D}F = 0$ is consistent with a symmetry such as the rotation symmetry in this example if the constraint commutes with the symmetry generators.

There is a constraint in a similar sense that one can impose on a superfield $\Phi$ which has the form of some differential operator $\bar{\mathbf{D}}$ such that $\bar{\mathbf{D}}\Phi = 0$. To satisfy supersymmetry, also $\bar{\mathbf{D}}(\Phi + \mathbf{Q}\Phi) = 0$ must be true, and assuming that $\bar{\mathbf{D}}$ is linear, also $\bar{\mathbf{D}}\mathbf{Q}\Phi = 0$ must be fulfilled. Superfields with $\bar{\mathbf{D}}\mathbf{Q}\Phi = 0$ are called chiral. The condition that $\bar{\mathbf{D}}$ and $\mathbf{Q}$ either commute or anticommute is a sufficient condition, and

$$\bar{\mathbf{D}}_\beta = \frac{\partial}{\partial \bar{\theta}} - i\sigma\theta\partial \qquad \mathbf{D}_\alpha = \frac{\partial}{\partial \theta} - i\bar{\theta}\sigma\partial$$

is the same as $\bar{\mathbf{Q}}_\beta$ and $\mathbf{Q}_\alpha$ in (3.5) except for the sign of the last term. The operators $\mathbf{D}$ and $\bar{\mathbf{D}}$ both anticommute with $\mathbf{Q}$ and $\bar{\mathbf{Q}}$. Constraints $\bar{\mathbf{D}}\Phi = 0$ and $\mathbf{D}\Phi = 0$ both together would mean $\Phi = 0$. Only $\bar{\mathbf{D}}\Phi = 0$ simplifies $\Phi(x^\mu, \theta, \bar{\theta})$ to $\Phi(x^\mu + i\bar{\theta}\sigma^\mu\theta, \theta)$ called a chiral superfield. With the substitution $y^\mu = x^\mu + i\bar{\theta}\sigma^\mu\theta$ this gives the simplest possible superfield $\Phi = \varphi + \bar{\psi}\theta + F\theta\theta$.

The action is $\int \Lambda(\theta, \bar{\theta}, x)\, d^4x\, d^2\theta\, d^2\bar{\theta}$ or $\int \mathcal{L}\, d^4x$ with $\mathcal{L} = \int \Lambda(\theta, \bar{\theta}, x)\, d^2\theta\, d^2\bar{\theta}$. Using the simple example $\Lambda = \Phi^*\Phi$ the result is an action of the form

$$\int \mathcal{L}\, d^4x = \int \left[\varphi^*(x)\frac{\partial^2\varphi(x)}{\partial x^2} + \frac{\partial\bar{\psi}_\alpha(x)}{\partial x^\mu}\sigma^\mu_{\alpha\beta}\psi_\beta(x) + F^*(x)F(x)\right] d^4x$$

after some calculations. The first term comes from a scalar boson, the second term shows a Dirac-like fermion and the last term is trivial because it does not contain derivatives. The Feynmann propagators contributing to this Lagrangian include a propagator for $\varphi$ and a propagator for $\psi$ as parallel lines plus a propagator of length zero for $F$ leading to a delta-function. This corresponds to a massless boson $\varphi$ and a massless fermion $\psi$ without interaction plus the trivial boson $F$.

The super-Lagrangian $\Lambda = \Phi^*\Phi$ in this example did not have any derivatives in it. The derivatives in the Lagrangian comes from the fact that the argument of the superfield had to be shifted. A super-Lagrangian $\Lambda = m\Phi^2$ gives a Lagrangian of the form $m(F\varphi + \bar{\psi}\bar{\psi})$ with a Feynman diagram where the fermionic term $m\bar{\psi}\bar{\psi}$ corresponds to a mass term and where the second-order diagram for the boson turns a $\varphi$ into an $F$ and back into a $\varphi$ with $m^2$ and is indistinguishable from a mass term $m^2\varphi^2$. An additional term $g\Phi^3$ in the super-Lagrangian gives two terms $3\,g\,\varphi^2 F$ and $3\,g\,\phi\bar{\psi}\bar{\psi}$ corresponding to two $\varphi$ joining into an $F$ and splitting again in two $\varphi$ with $g^2$ as well as an $\bar{\psi}$ emitting a $\varphi$ with $g$ (like an electron emitting a photon).

## 3.12 Spontaneous Breaking of Supersymmetry

Because supersymmetry is not an exact symmetry, there must be some kind of symmetry breaking. If there is a vacuum with zero energy, then the vacuum is supersymmetric. If a component of a superfield is not equal to zero, that breaks the symmetry, because this components get mixed up into each other under a supersymmetric transformation. If the vacuum has not zero energy, then this has no implications except for gravity, because this means that the vacuum gravitates. If the vacuum has an $F$-term and therefore has a non-zero component of the superfield pointing in some direction in superspace then the supersymmetry is broken. Working out the equations shows that $F$ has the effect of splitting the masses of different particles which rotate into each other under the supersymmetry transformation.

The main lesson to draw from this is that the theory of supersymmetry breaking is very close in spirit to conventional spontaneous symmetry breaking. It does not require the equations to be non-supersymmetric, but requires the solutions of the equations to be non-symmetric with respect to the supersymmetry. Another consequence is that when the vacuum is not supersymmetric, this gives something new. This is similar to the case when the vacuum is not invariant under rotation, because rotation of the vacuum gives in this case something other than the vacuum. The vacuum is formally counted as a boson state. If the vacuum is not symmetric under supersymmetry, transformation with respect to supersymmetry gives a fermion. Thus, the result is that whenever supersymmetry is spontaneously broken in this way, there is a massless particle, and the massless particle is a fermion. It is sometimes called a Goldstone fermion and sometimes called a Goldstino. This Goldstone fermion is the new thing that gets created when the transformation of the vacuum is not the vacuum.

Gravity has a significant effect here. If supersymmetry is broken a positive vacuum energy results in a cosmological constant which accelerates the expansion of the universe. The theory of supersymmetry is connected to a theory of supergravity which may ensure that the expansion is what has been measured.

## 3.13 Grand Unified Theories

The Standard Model is based on SU(3)×SU(2)×U(1) where SU(3) is the group of color and of quarks, SU(2) is the weak interactions and U(1) is closely related to the electromagnetic interactions. The group SU(3)×SU(2)×U(1) consists of separate transformations of three things, two things and one thing. The question is whether this group can be embedded into a group which is more symmetric and more simple, and indeed SU(3)×SU(2)×U(1) fits nicely into the group SU(5) which is the smallest group containing it as a subgroup. The SU(3)-multiplets contain the quarks, and SU(2)-multiplets contain the leptons. Thus, the hope is that quarks, leptons and other particles fit into something more general. There are several attempts for a grand unified theory but here only the one based on SU(5) is presented.

The group SU(5) consists of the $5 \times 5$ unitary matrices $\mathbf{U}$ with determinant $\det \mathbf{U} = 1$. The generators $\mathbf{G}$ are the small deviations from the identity $\mathbf{I} + i\varepsilon\mathbf{G} = \mathbf{U}$, and the group is characterized by the generators and their commutation relations. The generators are hermitian and traceless. There are 5 real diagonal elements and 10 complex off-diagonal elements giving 25 independent parameters in such $5 \times 5$ matrices, but it is reduced by one because of the condition that they are traceless. Thus, there are 24 generators.

The question is which subset of these 24 generators correspond to the generators of SU(3), SU(2) and U(1). A $5 \times 5$ matrix can be split such that the upper left $2 \times 2$ matrix is a generator of SU(2) and the lower right $3 \times 3$ matrix is a generator of SU(3). The three generators of SU(2) are the Pauli-matrices which are denoted here by $\tau_i$ in the upper left corner with zeros everywhere else. The eight generators of SU(3) are denoted $\lambda$ placed as $3 \times 3$ matrices in the lower right corner with zeros everywhere else. There is also a generator for U(1) which is a diagonal matrix with $-1$ in the upper two diagonal elements because of old conventions and $\frac{2}{3}$ in the lower three diagonal elements to become traceless. It is therefore a multiple of the identity in the upper left corner corresponding to SU(2) and in the lower right corner corresponding to SU(3). This generator is called $Y$ and distinguishes the SU(2)-block from the SU(3)-block.

Because there are 24 generators where the two W-bosons, the Z-boson and the photon of the electroweak interaction plus the gluons give 12, there must be 12 additional gauge bosons in this theory. They correspond to new forces. On the side of the fermions only the left-handed particles have weak interactions, and handedness changes when going form particle to antiparticle. Thus, instead of calling the particle and its antiparticle the independent objects, one can call the left-handed particle-antiparticle pair the independent objects. The five left-handed fermions in one family are the neutrino $\nu_e$, the electron $e^-$, and the antiquark down $\overline{d}$ in three colors. The remaining fermions are the down quark $d$, the up quark $u$, and the up antiquark $\overline{u}$ each in three colors plus the positron $e^+$, together they are ten particles. The antiquarks $\overline{u}$ and $\overline{d}$ have no SU(2)-properties because their antiparticles are right-handed quarks. This leads to

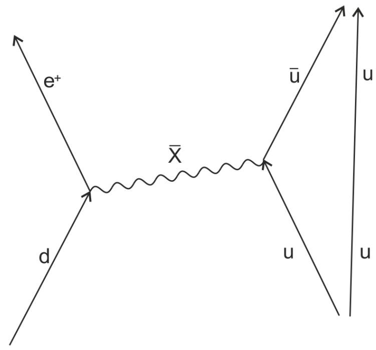
$$\begin{pmatrix} \mathrm{SU}(2) & & 0 & 0 & 0 \\ & & 0 & 0 & 0 \\ 0 & 0 & & & \\ 0 & 0 & & \mathrm{SU}(3) & \\ 0 & 0 & & & \end{pmatrix} \begin{pmatrix} \nu_e \\ e^- \\ \overline{d} \\ \overline{d} \\ \overline{d} \end{pmatrix}$$

with the particles selected this way to make sure that their total charge (the trace) is zero. There is a ten-dimensional representation of SU(5) with exactly the properties for the remaining ten particles of the first family. (Note that there is no left-handed antineutrino in the Standard Model.)

Before continuing with SU(5) some properties of SU($N$) are shown. The general SU($N$) group has among other representations a defining representation in form of $N \times N$ unitary matrices acting on vectors with $N$ elements and mixing their components. This representation is also called $N$. Taking the complex conjugate of the matrices and vectors gives a different representation of the same group which is called the complex conjugate representation or $\bar{N}$. If particles mix under $N$ then the antiparticles mix under $\bar{N}$. If one particle can be in one of $N$ states two particles can be in one of $N^2$ states. The corresponding representation $N \times N$ is reducible because there are the symmetric and the antisymmetric representations. The possible spins of two electrons, for example, have the three symmetric states $|uu\rangle$, $|dd\rangle$, $(1/\sqrt{2})(|ud\rangle + |du\rangle)$ corresponding to the total spin one and the antisymmetric singlet state $(1/\sqrt{2})(|ud\rangle - |du\rangle)$ corresponding to the total spin zero. Also the representation $N \times \bar{N}$ for a particle and an antiparticle can be split into an adjoint representation with $N^2 - 1$ traceless elements and one singlet as for the eight gluons of Quantum Chromodynamics in SU(3).

The SU(5) group mixes up the five left-handed particles $\nu_e$, $e^-$ plus $\overline{d}$ in the three colors which form a multiplet under the group SU(5). The transformations of SU(2) mix neutrino and electron mediated by W bosons, and the transformations of SU(3) mix the colors of the antiquark $\overline{d}$ mediated by gluons but SU(5) can also mix leptons and quarks mediated by new gauge bosons. When a $\overline{d}$ turns into a $e^-$, it emits an X-boson, and when a $\overline{d}$ turns into a $\nu_e$, it emits a Y-boson. These new bosons turn leptons into quarks or vice versa, have color, and can emit and absorb gluons.

The 5 are $\nu_e$, $e^-$ plus three $\overline{d}$, and the $\bar{5}$ contains therefore $\nu_e$, $e^+$ plus three $d$. The following table contains the antisymmetric $\bar{5}\times\bar{5}$ arranged such that the remaining ten left-handed states not in the 5 have the quantum numbers (electric charge) of the row and the column as if they were composite objects:




| | $\nu_e$ | $e^+$ | $d_\mathrm{r}$ | $d_\mathrm{g}$ | $d_\mathrm{b}$ |
|---|---|---|---|---|---|
| $\nu_e$ | $0$ | $e^+$ | $d_\mathrm{r}$ | $d_\mathrm{g}$ | $d_\mathrm{b}$ |
| $e^+$ | $-e^+$ | $0$ | $u_\mathrm{r}$ | $u_\mathrm{g}$ | $u_\mathrm{b}$ |
| $d_\mathrm{r}$ | $-d_\mathrm{r}$ | $-u_\mathrm{r}$ | $0$ | $\bar{u}_\mathrm{b}$ | $\bar{u}_\mathrm{g}$ |
| $d_\mathrm{g}$ | $-d_\mathrm{g}$ | $-u_\mathrm{g}$ | $-\bar{u}_\mathrm{b}$ | $0$ | $\bar{u}_\mathrm{r}$ |
| $d_\mathrm{b}$ | $-d_\mathrm{b}$ | $-u_\mathrm{b}$ | $-\bar{u}_\mathrm{g}$ | $-\bar{u}_\mathrm{r}$ | $0$ |

The diagonal contains zeros and the lower left triangle is the negative of the upper right triangle because of the antisymmetry. A $5\times5$ is called a 10, and a $\bar{5}\times\bar{5}$ is is called $\overline{10}$ because they both have ten entries. All the fifteen left-handed fermions and antifermions are contained either in the representation 5 or in the representation $\overline{10}$. This is the fermion content of the theory called SU(5) unification.

How the gauge bosons photon, W-bosons and Z-boson as well as the gluons act is know from the Standard Model. The question is what the new gauge bosons X and Y of the SU(5) unification theory do. One of the new processes is very surprising. A down-quark emits an anti-X-boson and turns into a positron. The anti-X-boson is absorbed by an up-quark which turns into an up-antiquark. Thus, as shown in the figure a proton consisting of a down quark and two up quarks can decay into a positron and either a $\Pi^0$ meson consisting of the $u$ and the $\bar{u}$ or, less likely, when the $u$ and the $\bar{u}$ annihilate into a photon. This proton decay is a rather dangerous process, and if protons would have the habit to do so, human beings would not be here to talk about proton decay. This proton decay is a consequence of pretty much any kind of attempt to unify the electromagnetic, the weak and the strong interaction into one single group which is not a product group. Thus, the question arises why the proton decay does not happen frequently.

A possible answer is the mass of the X-boson which goes into a propagator giving a probability amplitude. The question is therefore how big the mass of the X-boson has to be such that there are still protons from the time of the big bang. If half of the protons decay in $10^{15}$ years, then one of $10^{15}$ protons should decay within one year in average, but no proton decay has ever been detected. Half life of the proton is longer than $10^{33}$ years and the mass of an X-boson must be in the order of $10^{16}$ GeV. This is about three orders of magnitude smaller than the Planck mass.

There is another indication which make this number very tempting. It is the running of coupling constants mentioned above. The electromagnetic, weak and strong coupling constants are not numerically the same, but should be the same in a unified theory. SU(5) is not a symmetry of nature, even not approximately, because of the enormous difference of the mass of the X-boson compared with the mass of the photon, the W-bosons, the Z-boson and the gluons. If SU(5) would be an unbroken symmetry, the mass of all these gauge bosons would be the same because they are all part of the same multiplet. This means that the symmetry is spontaneously broken. The Higgs phenomenon give the photon and the W-boson different mass and broke the SU(2)×U(1) symmetry at the same time. Thus, there must be a similar phenomenon that splits the electron with the neutrino on one side from the three down-antiquarks on the other side and that breaks the symmetry between them drastically. This symmetry breaking would account for the big mass difference between the gauge bosons of the Standard Model and the X- and Y-boson.

There are some other pieces of evidence for the $10^{16}$ GeV. At high energy levels, the mass of the X- and Y-bosons can be considered to be zero. The question is at which point the world appears SU(5)-invariant. The coupling constants are not constant and depend on energy scales. Plotting the coupling constants of SU(3), SU(2) and U(1) as $1/g^2$ depending on $\log(E)$ gives straight lines crossing at the more or less same place which is the unification energy around $10^{16}$ GeV. Beyond that point, all the three coupling constants are the same, and below this energy, symmetry is broken.

The relation between grand unification and supersymmetry is not very tight, but if one adds all of the superparticles into the running of the coupling constants, then the three lines cross within 1% at the same place. That is one of the reasons why there is excitement about supersymmetry and grand unification.