GROUPS: NATURE'S "INVISIBLE HAND": AN ESSAY ON THE STANDARD MODEL

Gerald E. Marsh
gemarsh@uchicago.edu

Group theoretical methods did not receive a warm reception when introduced into the physics community. As put by John Slater in his 1975 autobiography, "Wigner, Hund, Heitler, and Weil entered the picture with their 'Gruppenpest' . . . The authors of the 'Gruppenpest wrote papers which were incomprehensible to those like me who had not studied group theory . . . The practical consequences appeared to be negligible, but everyone felt that to be in the mainstream one had to learn about it. . . .It was a frustrating experience, worthy of the name of a pest."

The "pest" was never vanquished. Today, group theory is fundamental to the Standard Model of particle physics and plays an important role in many other areas of physics as well. The introduction of groups into physics could be thought of as comparable to Adam Smith's introduction in his *Wealth of Nations* of the "Invisible Hand" into economics; with a little poetic license, groups could be said to play the role of Nature's Invisible Hand.

This essay will take the reader from some elementary ideas about groups to the essence of the Standard Model.

Introduction

It is the purpose of this essay to take the reader from some elementary ideas about groups to the essence of the Standard Model of particle physics along a relatively straight and intuitive path. Groups, from a pedagogical point of view, are usually introduced relatively late in a physics education. I will begin with them here to arrive at a semblance of the Dirac equation. This is followed by introducing the very essence of elementary quantum theory to obtain the actual Dirac equation, which governs the motion of the quarks and leptons of the Standard Model. An introduction to the gauge principle is then given and this will take us via the groups introduced in the beginning to an introduction to the Standard Model. In following this path, many technical details, and much of the physics, will be ignored. The idea is to give an Olympian view of this evolution, one that is often missing when absorbing the detailed subject matter of the Standard Model as presented in an historical approach to the subject.

Groups

In discussing the general mathematical idea of symmetry, Herman Weyl made the statement: "As far as I see, all *a priori* statements in physics have their origin in symmetry." But even our earliest perceptions of space and time and the invariance associated with them lead to the concept of groups.

The limited material introduced here from the vast field of group theory attempts to avoid the extensive background needed for a precise presentation. It also draws on how the material is often presented in the physics literature, which is often imprecise if not down right sloppy from a mathematical perspective. Even so, since most of the potential readers of this essay are expected to be from the physics community, it is important to make the connection to the physics literature. As much as possible, the notation is consistent with that used in physics.

When speaking of symmetry in quantum field theory one often defines different types of symmetry, the broadest division being into manifest symmetry, meaning the apparent type of symmetry found in the translation or rotation groups, and hidden symmetry where the symmetry only appears for special values of a parameter. This is the type that will be discussed below when symmetry breaking is introduced. There is a further distinction into local and global symmetries where local means that the parameters of the group depend on space-time location whereas global symmetries do not. Local symmetry, and its relation to dynamics is the foundation of all gauge theories.

Groups are abstract entities that are defined very broadly. They are required to satisfy the requirements that they have a closed binary operation that is associative, an identity element (also sometimes called a unit element), and each element must have an inverse. The closure property guarantees that the binary composition operation does not result in elements outside of the group. Group representations allows groups to act on vector spaces over fields such as the real or imaginary numbers. Groups can, and often do, have representations as matrices, and this is the representation that will be used here. For the cognoscenti, a matrix representation of a group G is a homomorphism (a mapping that preserves the group structure) from G onto $GL(n, R)$ or $GL(n, C)$.

Our focus will be on continuous groups (Lie groups) and we will begin with the simplest example of such a group, the set of all complex phase factors $\mathrm{U}(\theta) = \mathrm{e}^{i\theta}$. These phase factors form a unitary group called U(1), which when treated as a manifold (a Lie group) is 1-dimensional. Here unitary simply refers to all complex numbers with modulus unity.

Let us move on to two dimensions. If x_1 and x_2 are the coordinates of a point in a plane, we can transform these coordinates by use of a linear transformation represented by a real matrix; that is, if x is a one by two column matrix with entries x_1 and x_2, $x' = Ax$, or

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

If the $det(A) = a_{11}a_{22} - a_{12}a_{21} \neq 0$, and we require that length be preserved so that $x_1'^2 + x_2'^2 = x_1^2 + x_2^2$, there will be constraints on the elements of the matrix A. Simply substituting the transformation above into the latter expression for the requirement that the length be preserved results in the conditions:

$$a_{11}^2 + a_{21}^2 = 1,\ a_{11}a_{12} + a_{21}a_{22} = 0,\ a_{12}^2 + a_{22}^2 = 1,$$

which in turn imply that $\left(a_{11}a_{22} - a_{12}a_{21}\right)^2 = 1$, or $\det(A) = \pm 1$. The transformation is a rotation and the plus sign corresponds to a proper rotation and the minus to one that changes the orientation. We will restrict further consideration to only proper rotations where det(A) = +1. There are three conditions here imposed on four parameters leaving only one free parameter. This, of course corresponds to the angle of rotation in the plane. It is readily confirmed that these transformations form a group, which is known as the special orthogonal group in two dimensions, SO(2), and the term “special” corresponds to the choice of det(A) = +1. The requirement that the length be invariant can easily be extended to higher Euclidean dimensions to yield the groups SO(N).

A Hilbert space is a linear space over the field of complex numbers, meaning that if α and β are vectors in a Hilbert space then $\alpha + \beta$ is in the Hilbert space as is $c\alpha$, where c is any complex number. More precisely, a Hilbert space is an inner product space, which—as a metric space—is complete. In quantum mechanics the state of a system is a unit vector in Hilbert space. Any symmetry will then be connected to a unitary or anti-unitary transformation on this space (the unitary property will be discussed shortly).

Suppose we now allow the entries a_{ij} in the transformation matrix above to be complex numbers, and in addition require the transformation to have $|x_1|^2 + |x_2|^2$ as an invariant. As above, we now obtain the conditions

$$|a_{11}|^2 + |a_{21}|^2 = 1,\ |a_{12}|^2 + |a_{22}|^2 = 1,\ a_{11}a_{12}^* + a_{21}a_{22}^* = 0.$$

These conditions are equivalent to requiring $A^\dagger A = 1$, and the determinant of the matrix has modulus unity. Here † designates the transpose of the matrix and the complex conjugate of the elements. This is known as the Hermitian conjugate while * is the complex conjugate. Matrices satisfying these requirements belong to the representation of the unitary group U(2). If we now add the additional requirement that the determinant of the matrix is unity, this will result in $|a_{11}|^2 + |a_{12}|^2 = 1$, and the transformation matrix will have the special form

$$\begin{pmatrix} a_{11} & a_{12} \\ -a_{12}^{*} & a_{11}^{*} \end{pmatrix}.$$

From the original eight free parameters there are now only three. These matrices are known as the special unitary matrices for two dimensions or SU(2). Special unitary transformations are especially important in quantum mechanics and for what follows. Higher dimensional special unitary groups may also be defined and are known as SU(N), and SU(3) will play an important role later.

Two other concepts from group theory will be relevant in what follows, that of a normal subgroup and a factor group. If one has a group G and an element g where $g \in G$, and a subgroup $N \subset G$, if N is a normal subgroup then Ng, the set of all elements of N multiplied by $g \in G$ on the right, is the same as the set of all elements of N multiplied by g on the left; that is $gN = Ng$ or $gN - Ng = 0$. Another way of writing this is $gNg^{-1} = N$, which means that N is left invariant by every inner automorphism of G. When this is the case, N is said to be self-conjugate. If a group contains normal subgroups then it may be expressed as being made up of smaller groups. The expression gN is known as a coset and when N is normal the cosets themselves form a group known as a factor (or quotient) group written G/N. One says that the group of cosets of N under the induced operation (taken from G) is the factor group of G modulo N. Although it is often said that the cosets are residue classes of G modulo N, it might be better to say that the left cosets are residue classes of the group homomorphism $f{:}G \to G/N$ defined by $g \to gN$, and similarly for the right cosets. Groups having no non-trivial normal subgroups are known as simple groups; that is, a simple group G has only the identity and G as normal subgroups. The Standard Model of particle physics is made up of products of simple groups.

Matrix representations were introduced above as a homomorphism from a group G onto GL(n, R) or GL(n, C); these groups operate on a vector space V over the real or complex numbers. A reducible representation is one where the vector space contains an invariant subspace—one that gets mapped onto itself; if the representation contains no such invariant subspace it is called irreducible. Irreducible representations are the building blocks for all finite-dimensional completely reducible representations.

Here are a few additional facts about normal (invariant) subgroups and their mappings:

Suppose that N is a normal subgroup of G. Then there is a bijective—one-to-one and onto—mapping between irreducible representations of G/N and irreducible representations of G having N in the kernel; i.e., N is mapped onto the identity element.

The direct product of the groups G and N is written as $G \times N$. If N is an invariant subgroup of G, then the group G is the direct product of the invariant subgroup N with the factor group G/N; that is, $G = (G/N) \times N$.

A group G is the direct product of its subgroups, say N_1 and N_2, if N_1 and N_2 are normal subgroups that are disjoint, that is, $N_1 \cap N_2$ = Identity, and they generate the group so that $G = N_1 N_2$, where $N_1 N_2 = \{ n_1 n_2 \mid n_1 \in N_1, n_2 \in N_2\}$.

A Semblance of the Dirac Equation from Groups

In quantum mechanics particles are characterized by, in addition to their positions in space and time: transformations which require the introduction of the Lorentz group of special relativity; electric charge; and transformations of any internal degrees of freedom such as spin. To avoid the unnecessary complications that arise from the introduction of the inhomogeneous Lorentz, or Poincare group (which generally includes translations), the discussion to follow will be to a large extent restricted to the proper orthochronous Lorentz group.

Eugene Wigner's contributions to the role of groups in physics are responsible for groups playing such an important role in the Standard Model. In discussing Wigner's classification of the irreducible representations of the Poincaré group, the connection between groups and elementary particle has been succinctly stated by Sternberg: "An elementary particle 'is' an irreducible unitary representation of the group, G of physics, where these representations are required to satisfy certain physically reasonable restrictions . . . ."

It is assumed in what follows that the reader is familiar with Minkowski space and special relativity, as well as the Dirac equation and its usual derivation and solutions. Our purpose here is to show that a semblance of the Dirac equation can be derived using only the properties of groups and special relativity. The equation corresponds to the relationship between the two spinors that come from the representations (½,0) and (0, ½) of the Lorentz group. Those not familiar with the spinor representations of the Lorentz group are directed to Appendix I.

The discussion below is based on that given in Ryder's book on quantum field theory listed in the bibliography.

Under a general Lorentz transformation, there are two types of 2-component spinors. They correspond to the right-handed, (½, 0), or the left-handed, (0, ½), representations of the Lorentz group. These are two-dimensional representations that are interchanged by Hermitian conjugation. The 2-spinors associated with each representation can be put together into a single 4-spinor with two components labeled Φ_R and Φ_L designating right and left helicity, which is defined as the component of spin in the direction of the momentum. With the constraints given above, the 4-spinor would transform under a pure Lorentz transformation without rotation (proper Lorentz transformations include a boost and a rotation, which when composing non-colinear boosts leads to the Thomas precession) as

$$\Phi' = \begin{pmatrix} \Phi_R \\ \Phi_L \end{pmatrix}' = \begin{pmatrix} e^{\frac{1}{2}\boldsymbol{\sigma}\cdot\boldsymbol{\phi}} & 0 \\ 0 & e^{-\frac{1}{2}\boldsymbol{\sigma}\cdot\boldsymbol{\phi}} \end{pmatrix} \begin{pmatrix} \Phi_R \\ \Phi_L \end{pmatrix},$$

where $\boldsymbol{\sigma}$ are the 2×2 Pauli matrices and $\boldsymbol{\phi}$ is the hyperbolic angle for the Lorentz transformation; i.e., since $\gamma^2 - \beta^2\gamma^2 = 1$, one may set $\gamma = \cosh\phi$, $\gamma\beta = \sinh\phi$, where $\gamma = (1-v^2/c^2)^{-1/2}$ and $\beta = v/c$. The reader is referred to Appendix I or the group theory literature to fully understand how the exponential forms in the transformation matrix arise.

Consider first Φ_R. If the exponent of the relevant exponential is written as $-(i/2)\ \boldsymbol{\sigma}\bullet(i\boldsymbol{\phi})$ and the exponential expanded one obtains

$$\Phi_R \rightarrow e^{\frac{1}{2}\boldsymbol{\sigma}\bullet\boldsymbol{\phi}}\Phi_R = \left[\cosh\left(\frac{\phi}{2}\right) + \boldsymbol{\sigma}\bullet\hat{\boldsymbol{n}}\ \sinh\left(\frac{\phi}{2}\right)\right]\Phi_R,$$

where $\hat{\boldsymbol{n}}$ is a unit vector in the direction of the Lorentz boost. Now assume initially that $\Phi_R(0)$ is transformed by the boost to $\Phi_R(\boldsymbol{p})$; i.e., that the particle is initially at rest with momentum zero and the boost gives it momentum $\boldsymbol{p}$. Then from the hyperbolic half angle relations and the definition of $\cosh\phi$ and $\sinh\phi$ given in the paragraph above, as well as the fact that in units where $c = 1$ the total energy E of a particle with momentum $\boldsymbol{p}$ is $E = \gamma\, m$, and noting that $\boldsymbol{p}$ is in the same direction as $\hat{\boldsymbol{n}}$, we obtain

$$\Phi_R(\boldsymbol{p}) = \frac{E + m + \sigma\bullet\boldsymbol{p}}{\left[2m(E+m)\right]^{\frac{1}{2}}}\ \Phi_R(0).$$

Similarly, for $\Phi_L \rightarrow e^{-\frac{1}{2}\boldsymbol{\sigma}\bullet\boldsymbol{\phi}}\Phi_L$,

$$\Phi_L(\boldsymbol{p}) = \frac{E + m - \sigma\bullet\boldsymbol{p}}{\left[2m(E+m)\right]^{\frac{1}{2}}}\ \Phi_L(0).$$

For $\Phi(0)$, the distinction between left and right becomes meaningless since for $\boldsymbol{p} = 0$ the there is no direction of momentum to which the spin can be aligned. Consequently, one can set $\Phi_L(0) = \Phi_R(0)$. Using this and the definition of the four momentum, $p_\mu = (\mathrm{E}, -\boldsymbol{p})$, the last two equations may be expressed in matrix form as

$$\begin{pmatrix} -m & p_0 + \sigma\bullet\boldsymbol{p} \\ p_0 - \sigma\bullet\boldsymbol{p} & -m \end{pmatrix}\begin{pmatrix} \Phi_R(\boldsymbol{p}) \\ \Phi_L(\boldsymbol{p}) \end{pmatrix} = 0.$$

Remembering that this is a four dimensional matrix (Φ_R, Φ_L, each have two components while $\boldsymbol{\sigma}$, and $\boldsymbol{p}$ have three) one may define the four dimensional matrices

$$\psi(p) = \begin{pmatrix} \Phi_R(\boldsymbol{p}) \\ \Phi_L(\boldsymbol{p}) \end{pmatrix}, \quad \gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & -\sigma^i \\ \sigma^i & 0 \end{pmatrix},$$

where 1 corresponds to the two by two identity matrix. With these definitions, the previous matrix equation may be written as

$$\left(\gamma^0 p_0 + \gamma^i p_i - m\right)\psi(p) = 0.$$

This is the promised semblance of the Dirac equation. It was derived using only the properties of groups and special relativity. The equation obtained corresponds to the relationship between the two spinors that come from the representations (1/2,0) and (0,1/2) of the Lorentz group. To get the actual Dirac equation one must introduce some minimal elements of quantum mechanics.

It should be noted here that the derivation above uses the chiral representation, so called because Φ_R and Φ_L are eigenstates of chirality (the term "chirality" being equivalent to "handedness"). In the standard representation, generally used to study the Dirac equation, the definitions of the γ-matrices are different.

Minimalist Quantum Mechanics

The origination of quantum mechanics dates back to Max Planck in 1900 and his studies of heat radiation that led him to introduce the postulate that energy came in discrete, finite quanta of energy $h\nu$. Planck was awarded the 1918 Nobel Prize in Physics for his work but was never comfortable with the idea of quanta. Nonetheless, essentially all of quantum theory follows from special relativity and Planck's discovery that $E = h\nu$. We will use this fact and de Broglie's discovery of the wave nature of matter to obtain what is required to convert the "almost Dirac equation" to the quantum mechanical version.

De Broglie in his 1924 publication "*Recherches sur la Théorie des Quanta*" introduced the thesis that elementary particles had associated with them a wave, what we call the wave function, and what de Broglie called an "*onde de phase*" or a phase wave. It is a consequence of the relation $E = h\nu$. In his 1929 Nobel lecture he used the following argument:

$$p = \gamma m v = \gamma m c^2 \frac{v}{c^2} = \mathrm{E}\,\frac{v}{c^2}.$$

He now identifies the energy E of a *massive* particle with $E = h\nu$ to give

$$p = \frac{h\nu}{c^2/v}.$$

This identification is the key step used by de Broglie in deriving his relation. Since the velocity of the massive particle is always less than that of light, so that $c^2/v > c$, he states that "*qu'il ne saurait être question d'une onde transportant de l'énenergie*" (it is not a question of a wave transporting energy). Consequently, he makes another key assumption that c^2/v corresponds to a phase velocity via $vv_{ph} = c^2$, so that

$$p = \frac{h}{v_{\mathrm{ph}}/\nu}.$$

Since $v_{ph} = \nu\lambda$, de Broglie obtains his fundamental relation $\lambda\, p = \mathrm{h}$.

The two relations, $E = h\nu$ and $\lambda\, p = \mathrm{h}$ allow us to derive what is needed to transform the "semblance of the Dirac equation" given above to the quantum mechanical version of the Dirac equation. Start with a classical wave packet propagating in the $\boldsymbol{k}$ direction,

$$\Psi(\boldsymbol{r},t) = \int \mathrm{F}(\boldsymbol{k})\, \mathrm{e}^{i(\boldsymbol{k}\bullet\boldsymbol{r} - \omega t)} \mathrm{d}k.$$

Now use the two quantum mechanical relations we have to transform this to

$$\Psi(\boldsymbol{r},t) = \int \mathrm{F}(\boldsymbol{p})\, \mathrm{e}^{i(\boldsymbol{p}\bullet\boldsymbol{r} - Et)/\hbar} \mathrm{d}p.$$

Take the time derivative to get one expression and the gradient to get a second:

$$\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t} = -\frac{i}{\hbar}\int \mathrm{E}\,\mathrm{F}(\boldsymbol{p})\, \mathrm{e}^{i(\boldsymbol{p}\bullet\boldsymbol{r} - Et)/\hbar} \mathrm{d}p,$$

$$\nabla\Psi(\boldsymbol{r},t) = \frac{i}{\hbar}\int \boldsymbol{p}\,\mathrm{F}(\boldsymbol{p})\, \mathrm{e}^{i(\boldsymbol{p}\bullet\boldsymbol{r} - Et)/\hbar} \mathrm{d}p.$$

The first of these implies that $\mathrm{E} = i\hbar\frac{\partial}{\partial t}$ and the second that $\boldsymbol{p} = -i\hbar\nabla$. By substituting these expressions for E and $\boldsymbol{p}$ into the expression for the 4-momentum $p_\mu = (\mathrm{E}, -\boldsymbol{p})$ the semblance of the Dirac equation derived above from special relativity and group theory alone becomes the actual, quantum mechanical Dirac equation.

Gauge Principle

The gauge principle is fundamental to the standard model of particle physics. All of the observed interactions of elementary particles and their associated quanta are a result of vector fields created by the transition from global to local gauge symmetries.

Let us begin with the electromagnetic field. It is well known today—for example from the Aharanov and Bohm experiment, that in the presence of an electromagnetic field the wave function of a charged particle acquires a phase factor that depends on the vector potential A_μ (4-vector)

$$\Psi(\boldsymbol{x}, t) \rightarrow \Psi'(\boldsymbol{x}, t) = \mathrm{e}^{-\frac{ie}{\hbar}\int A_\mu dx^\mu}\Psi(\boldsymbol{x}, t).$$

It is also true that this transformation is equivalent to what is known as "minimal coupling"—the replacement in the free particle Lagrangian of the partial derivatives by the "gauge-covariant" derivative, $\partial_\mu \rightarrow D_\mu = \partial_\mu - ieA_\mu$.

Writing the phase factor more generally as $\mathrm{e}^{-ie\phi(\boldsymbol{x},t)}$, if $\phi(\boldsymbol{x}, t)$ *is* a constant, then $\Psi'(\boldsymbol{x}, t)$ *is* a solution of the free-particle wave equation. This corresponds to a *global* phase invariance since ϕ is constant throughout space and time. Allowing ϕ to have a space-time dependence means we must introduce the electromagnetic field derivable from the 4-potential A_μ. Put another way, if ∂_μ is replaced by $D_\mu = \partial_\mu - ieA_\mu$ in the free-particle wave equation, the resulting wave equation can be made gauge invariant if both of the equations

$$\begin{aligned}
&\Psi(\boldsymbol{x}, t) \rightarrow \Psi'(\boldsymbol{x}, t) = \mathrm{e}^{-ie\phi(\boldsymbol{x},t)}\Psi(\boldsymbol{x}, t) \\
&A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\phi(\boldsymbol{x}, t)
\end{aligned}$$

are satisfied, thereby introducing a *local* phase invariance. The 4-vector A_μ' that results from allowing ϕ to have a space-time dependence lets the vector potential be considered to be a gauge field, and the fact that the interaction is determined by this field is known as the gauge principle. This principle also holds for relativistic equations like the free-particle Dirac equation

$$\left(\gamma^\mu\partial_\mu + m\right)\psi(x) = 0.$$

When the transformation $\partial_\mu \rightarrow D_\mu$ is made to this equation it introduces the electromagnetic interaction, which, when radiative corrections are included, has been found to be correct to many decimal places.

At this point, in order to generalize the covariant derivative for fields other than electromagnetism, we need to introduce the idea that continuous groups have infinitesimal group generators. Above we saw that groups have parameters such as the angles of rotation. If all of the parameters are set equal to zero, the group matrix representation becomes the identity element of the group. For sufficiently small values of, say, k parameters α_i, an element of the group G(α) lying close to the identity may be represented as a Taylor series expansion

$$\mathrm{G}(\alpha) = \mathrm{G}(0) + \sum_{i=1}^{k} \alpha_i \left(\frac{\partial \mathrm{G}}{\partial \alpha_i}\right)_{\alpha_i = 0} + \ldots$$

where G(0) is the identity matrix. The infinitesimal group generators, X_i are

$$X_i = \left(\frac{\partial \mathrm{G}}{\partial \alpha_i}\right)_{\alpha_i = 0}.$$

Given the group generators, the vector potential A_μ of electromagnetism associated with the group U(1) may be generalized to other groups such as SU(2) and SU(3), by using the generators of these groups to write

$$A_\mu(x) = A_\mu^i(x) X_i,$$

where the $A_\mu^i(x)$ are ordinary vector fields and the Einstein summation convention is used here and in what follows. Here there is a sum over the repeated index i so as to include all generators of the group. The explicit matrix form for the generalized vector potential will be discussed below.

The symmetry transformations that result from the group generators—the first two terms of the expansion above with α_i taken to be very small—are the infinitesimal part of what is known as a Lie group. It can be shown that the generators obey the commutation relations

$$\left[X_i, X_j\right] = iC_{ij}^l X_l,$$

where the C_{ij}^{l} are a set of real constants called the structure constants, which depend on the particular group. Such sets of commutation relations comprise a Lie algebra. Note that the anti-symmetry of the commutator implies that the C_{ij}^{l} are also antisymmeteric in the lower indices i and j.

We will now need to find the form of the derivative D_μ for non-Abelian groups (where matrices of the representation do not commute).

Arbitrary gauge group local symmetry transformations and corresponding gauge covariant derivatives

Thus far we have the following ingredients for a gauge theory: A Lie group G(x) having an independent copy assigned to each point x of Minkowski space. In the mathematical literature, this type of structure is known as a fiber bundle. The relationship between the group assigned to each space-time point and the generators of the group, X_i, is

$$\mathrm{G}(x) = \mathrm{e}^{-ie\alpha^i(x)X_i}.$$

The space-time dependence is carried by the parameters $\alpha^i(x)$. As the particle moves through a potential field, the group G(x) tells us how a set of basis vectors in the "internal space" associated with the particle changes. This internal space corresponds to internal degrees of freedom associated with the particle such as isotopic spin.

The wave function of the particle can then be written as

$$\psi(x) = \psi_i(x)u^i,$$

where the sum is over the set of internal space basis vectors u^i. The index i identifies components of an internal space quantity such as isotopic spin; $\psi_i(x)$ should be thought of as a component of $\psi(x)$ in the basis u^i. For an infinitesimal displacement in space-time,

$$\mathrm{G}(\mathrm{d}x) = \mathrm{e}^{-iq\mathrm{d}\alpha^i X_i},$$

where $\mathrm{d}\alpha^i = \left(\partial_\mu \alpha^i(x)\right)\mathrm{d}x^\mu$. q is the electric charge e for electromagnetic gauge group U(1) or the coupling constant for an arbitrary gauge group. As the particle moves from x to x +dx, the wave function changes by

$$\mathrm{d}\psi(x) = \left(\partial_\mu \psi_i(x)\right)\mathrm{d}x^\mu u^i + \psi_i(x)\mathrm{d}u^i.$$

It is the second term on the right hand side of this equation that describes the change in the basis vectors; the first is the change in $\psi_\iota(x)$ in moving from x to x +dx. Now the generators X_i are matrices that act on the column vectors u^i of the basis. This means that G(dx)u^i can be written as

$$\mathrm{G}(\mathrm{d}x)u^i = \mathrm{e}^{-iq\left(\partial_\mu \alpha^k(x)\right)\mathrm{d}x^\mu (X_k)_{ij}}\, u^j.$$

Remembering from the commutator discussion above that the matrices representing the group generators must be antisymmetric, this can be expanded to first order in dx as

$$u^i + \mathrm{d}u^i = \left[\delta^i_j - iq\left(\partial_\mu \alpha^k(x)\right)\mathrm{d}x^\mu (X_k)_{ij}\right]u^j.$$

For $i = j$, when the group generator matrix vanishes, δ^i_j becomes the identity matrix and it is the only remaining term, vanishing otherwise. The second term within the brackets corresponds to du^i and allows the introduction of the generalized vector potential,

$$\left(A_\mu\right)_{ij} = \left(\partial_\mu \alpha^k(x)\right)\left(X_k\right)_{ij}.$$

The change in the wave function then becomes

$$\mathrm{d}\psi(x) = \left[\left(\partial_\mu \psi_\iota(x)\right)\delta_{ij} - \mathrm{i}q\left(A_\mu\right)_{ij}\psi_\iota(x)\right]\mathrm{d}x^\mu u^j.$$

δ_{ij} has been introduced so as to allow the u^i that appears in the first term within the brackets to be factored out. Inspection of this equation tells us that, if we introduce the gauge covariant derivative

$$\mathrm{D}_\mu \psi_j(x) = \left[\delta_{ij}\partial_\mu - \mathrm{i}q\left(A_\mu\right)_{ij}\right]\psi_\iota(x),$$

it can be written as

$$\mathrm{d}\psi(x) = \left(\mathrm{D}_\mu \psi_j(x)\right)\mathrm{d}x^\mu u^j.$$

The last few equations have been written out in great detail. They can be simplified considerably if the explicit matrix indices are suppressed so that we would have instead the relations as they are usually found in the literature:

$$A_\mu = A_\mu^k X_k,$$
$$\mathrm{d}\psi(x) = \left[\left(\partial_\mu \psi(x)\right) - \mathrm{i}qA_\mu \psi(x)\right]\mathrm{d}x^\mu = \left(\mathrm{D}_\mu \psi(x)\right)\mathrm{d}x^\mu,$$
$$\mathrm{D}_\mu = \partial_\mu - \mathrm{i}qA_\mu.$$

Standard Model Beginnings

The Standard Model brings together three of the fundamental forces of nature. It is a gauge theory of strong and electroweak interactions. This means that at each point of space-time there is an internal space attached. From the mathematical point of view, the structure is that of a fiber bundle with a Minkowski base space and a principal bundle consisting of the gauge group. Individually, the groups involved are SU(3) for the strong force, SU(2) for the weak force and U(1) for the electromagnetic force. The symmetries involved with these groups are internal symmetries of the internal space, the others being space-time symmetries of the kind discussed above. These groups are usually written with subscripts *C*, *L*, *Y*, and *EM* respectively, standing for color, left, hypercharge, and electromagnetic. The restriction *L* reflects the fact that nature does not seem to have right handed neutrino components.

When these groups are put together to represent the Standard Model, an element *g* contained in the combination can be written

$$g \in \mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y},$$

where $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ is the Glashow-Weinberg-Salam electroweak symmetry group and $\mathrm{U}(1)_\mathrm{Y}$ is the phase group of weak hypercharge. This symmetry can be "spontaneously" broken (to be discussed later) to U(1), the phase group of the usual electric charge. The symbol $\times$ means the direct product so that if $g \in \mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$, one may represent *g* as a 6 × 6 block diagonal matrix having the form

$$g = \begin{pmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{pmatrix}, \quad g_1 \in \mathrm{SU}(3)_\mathrm{C}, \ g_2 \in \mathrm{SU}(2)_\mathrm{L}, \ g_3 \in \mathrm{U}(1)_\mathrm{Y}.$$

g_1, g_2, and g_3 are elements contained in the simple groups $\mathrm{SU}(3)_\mathrm{C}$, $\mathrm{SU}(2)_\mathrm{L}$, and $\mathrm{U}(1)_\mathrm{Y}$, which have no non-trivial normal subgroups.

Some language: The local *weak isospin* symmetry $SU(2)_L$ governs the weak interactions between quarks and leptons, while $SU(3)_C$ governs the strong color interactions between quarks. Being spin ½ particles, quarks obey a form of the Dirac equation. Weak isospin should not be confused with the isotopic spin (or isospin) used by Heisenberg to describe the symmetry between the neutron and the proton, which would transform into one another under the spin ½ representation of SU(2). There is, however, a close relation between Heisenberg's isospin and weak isospin in that a nucleon's isospin is the sum of the weak isospins of its constituent three quarks.

Isospin

If strong hadronic forces are charge independent, an isospin vector, $\boldsymbol{I}$, can point in any direction in isotopic space. In the case of the neutron and proton, which are by definition distinguished by I^3, the charge operator Q corresponding to the electric charge q is $Q = e(I^3 + ½)$. Since the nucleons have spin ½, they have $2I + 1 = 2$ possible orientations in isotopic space, so that I^3 has the value of ½ or −½, and $e(I^3 + ½)$ has the value 0 or e.

In the Standard Model, members of the particle zoo are grouped into isospin multiplets where each member of the multiplet is identified with different orientation in isospin space in the same way as was done for the proton and neutron. Charged current experiments show that the leptons and associated neutrinos must be represented as left-handed "doublets" of isotopic spin so that for three generations one has

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L.$$

The π-meson, which has $I = 1$ so that there are $2I + 1 = 3$ members of the isospin multiplet, is consequently a triplet, corresponding to

$$I^3 = \begin{cases} +1 & \pi^+ \\ 0 & \pi^0 \\ -1 & \pi^- \end{cases}.$$

Here, the relationship between the charge and I^3 is $q = e\, I^3$. Gell-Mann and Nishijima generalized this relation to include the strange particles by assuming that the charge of other particles are related to I^3 by a linear relation of the form $q = a\,I^3 + b$. The constant a is determined from $q = e\,I^3$ for pions as $a = e$. b can be found as follows: I^3 has the range $-I$ to $+I$, giving the average charge of the multiplet as $\langle q \rangle = b$. Only particles with zero hypercharge have $\langle q \rangle = 0$; otherwise, $\langle q \rangle = (½)eY$, where Y is the hypercharge.† The generalization is then $q = e\,(I^3 + Y/2)$. Expanded to include other quantum numbers (originally only baryon number and strangeness) in addition to hypercharge, this is known as the Gell-Mann Nishijima relation. In terms of operators, the weak hypercharge Y is defined by

$$Q = T^3 + \frac{Y}{2},$$

T^3 being an element of the $SU(2)_L$ Lie algebra defined above, and Q being the charge operator that generates $U(1)_{em}$.

First, consider $SU(2)_L \times U(1)_Y$, the groups associated with electroweak unification; the group $SU(3)_C$ will be discussed when the fundamental particles of the Standard Model are introduced.

The electro-weak group $SU(2)_L \times U(1)_Y$

The charge operator Q is associated with $U(1)_{em}$, and similarly, the hypercharge operator Y is associated with $U(1)_Y$. As indicated by the names, weak isospin and weak hypercharge come from the Gell-Mann and Nishijima approach to forming SU(2) hadronic isospin multiplets.

Some terminology: Taken together, mesons and baryons are known as hadrons. While mesons are made up of quark-antiquark pairs, baryons are composed of three quarks. Mesons have baryon number zero, which is why they are composed of a quark and

† Hypercharge is defined as $Y = B + S$, where B is the baryon number and S is the strangeness. Later in this essay it will be seen that baryons are composed of three quarks, u, d, and s, so that the baryon number of quarks is 1/3 (antiquarks, −1/3). Strangeness counts the number of strange quarks or antiquarks comprising the states that make up a particle's wavefunction; e.g., the wavefunction for the K^0 meson, $\left| -d\bar{s} \right\rangle$, has a strangeness of +1, while that for the K^-, $\left| -s\bar{u} \right\rangle$, has strangeness −1.

antiquark pair having baryon number 1/3 and −1/3 respectively. Protons and neutrons, known as nucleons, have an attracting force acting between them that is due to residual color interactions. Only about one percent of the rest mass of these nucleons is due to their constituent quark masses—the rest is due to quark gluon interactions. The meaning of this will become clear in what follows.

The idea of using the groups $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ was introduced by Glashow in 1961 and predated the discovery of weak neutral currents. Experiments indicate that in addition to the photon field A_μ, the weak interactions require three intermediate vector bosons mediating processes such as the scattering of ν_μ by e^- and beta decay. Glashow's work was extended by Weinberg in 1967 and Salam in 1968 to include the required bosons, the $W^\pm$, defined below, and the Z^0. These gauge particles are massless and are given masses by means of the introduction of a scalar field called the Higgs field, which results in the spontaneous symmetry breaking that, except for the photon and neutrino, give the particles mass.

In order to achieve the unification under the Glashow-Weinberg-Salam model of the weak and electromagnetic forces a new weak neutral current interaction mediated by the Z^0 was introduced for reasons having to do with gauge invariance requirements related to the interaction between the W^+ with the W^- and the fact that the generators of the W^+ and W^- do not commute. The existence of the weak neutral current means that there is a weak force between electrons in addition to the Coulomb force so that Coulomb's law must be modified.

Including the new weak force in addition to the Coulomb force means that the usual vector potential A_μ of the U(1) group must be modified to be a linear combination of the U(1) gauge field and the new W_μ^3 field of SU(2). The Standard Model uses the resulting isotriplet of vector fields W_μ^i coupled with strength g to the weak isospin current J_μ^{i}, along with a vector field B_μ coupled to the weak hypercharge current j_μ^Y with strength conventionally taken to be $g'/2$.

There are two *observed* currents, the electromagnetic current j_μ^{em} and the neutral current J_μ^{NC}. These will be expressed in terms of the two neutral currents J_μ^3 and j_μ^Y belonging respectively to the symmetry groups $SU(2)_L$ and $U(1)_Y$. The following will indicate how this is done.

We now assume that both charged and neutral currents exist, that the charged currents only couple between left-handed leptons, and that the bosons mediating the weak interaction are the $W^\pm$ and Z^0, which are massless at this point.

Several fields will be introduced in what follows: $\boldsymbol{T} = \left(T^1, T^2, T^3\right)$ are the generators of $SU(2)_L$, Y the generator of $U(1)_Y$, and these generators obey the commutation relations

$$\left[T^i, T^j\right] = i\varepsilon_{ijk} T^k, \quad \left[T^i, Y\right] = 0.$$

W_μ^i, with $i = 1, 2, 3$ is an isospin triplet—meaning it is an isospin vector and a 4-vector in space-time, which couples to the weak isospin current J_μ^i, while B_μ is an isospin singlet—meaning it is an isoscalar and a 4-vector in space-time that couples to the weak hypercharge current j_μ^Y. These fields will be used to form the physical particles Z^0, W^+ and W^-, the last two being defined as

$$W^+ = \frac{1}{\sqrt{2}}\left(W_\mu^1 - iW_\mu^2\right), \quad W^- = \frac{1}{\sqrt{2}}\left(W_\mu^1 + iW_\mu^2\right).$$

The W^+ and W^- fields are charged bosons, while W_μ^3 and B_μ are neutral fields.

The leptons are left-handed doublets with isospin set equal to ½ and $T^3 = \pm$ ½ and right-handed singlets having zero isospin:

$$\begin{pmatrix} \psi_{\nu l} \\ \psi_l \end{pmatrix}_{\mathrm{L}} = \frac{1-\gamma_5}{2} \begin{pmatrix} \psi_{\nu l} \\ \psi_l \end{pmatrix}, \quad l = e, \mu, \tau$$

$$\left(\psi_l\right)_R = \frac{1+\gamma_5}{2}\ \psi_l,$$

where, for example, $\Psi_{\nu e}$ corresponds to the electron neutrino wavefunction.

The Pauli matrices,

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

may be used to define the step up and step down operators

$$\tau_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \tau_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

which are used to raise and lower the isotopic spin.

Using doublet $\begin{pmatrix} \psi_{\nu l} \\ \psi_l \end{pmatrix}_L$ introduced above, one may introduce an isospin triplet of weak currents,

$$J_\mu^i(x) = \frac{1}{2}\overline{\begin{pmatrix} \psi_{\nu l} \\ \psi_l \end{pmatrix}}_L \gamma_\mu \tau_i \begin{pmatrix} \psi_{\nu l} \\ \psi_l \end{pmatrix}_L, \quad i = 1, 2, 3.$$

The corresponding charges,

$$T^i = \int J_0^i(x)\, d^3x,$$

generate an $SU(2)_L$ Lie algebra,

$$\left[T^i, T^j\right] = i\varepsilon_{ijk} T^k.$$

The weak current $J_\mu^3(x)$ introduced above cannot be identified with the *experimentally observed* weak neutral current J_μ^{NC} because the latter has a right-handed component. The electromagnetic current is a neutral current with both left and right handed components, and is given by,

$$\mathrm{e} j_\mu^{em} = e\bar{\psi}\gamma_\mu Q\psi,$$

where Q is the charge operator having eigenvalue -1 for the electron. The left-handed component belongs to an isotriplet and will be associated with T^3 and W_μ^3 below, while the right-handed component is an isosinglet current that has both right and left-handed components. Neither j_μ^{em} or J_μ^{NC} obey SU(2)$_L$ symmetry. The idea now is to form two orthogonal combinations of $J_\mu^3(x)$ and the weak isospin singlet j_μ^Y, that have appropriate transformation properties under SU(2)$_L$. j_μ^Y is the weak hypercharge current given by

$$j_\mu^Y = \bar{\psi}\gamma_\mu Y\psi,$$

which is unaffected by SU(2)$_L$ transformations.

The isospin doublet and singlet introduced above are now required to be invariant under the local gauge transformations so that

$$\begin{pmatrix}\psi_{\nu l}\\ \psi_l\end{pmatrix}_L \rightarrow \begin{pmatrix}\psi_{\nu l}\\ \psi_l\end{pmatrix}_L' = e^{i\boldsymbol{\alpha}(x)\bullet \boldsymbol{T}}\begin{pmatrix}\psi_{\nu l}\\ \psi_l\end{pmatrix} = e^{i\boldsymbol{\alpha}(x)\bullet\frac{\boldsymbol{\tau}}{2}}\begin{pmatrix}\psi_{\nu l}\\ \psi_l\end{pmatrix},$$
$$(\psi_l)_R \rightarrow (\psi_l)_R' = e^{i\alpha(x)Y}(\psi_l)_R.$$

Note that $T^i = \tau^i/2$ and that the operator Y, which generates the U(1) group, is simply a constant Y. This gauge invariance will only hold if the Lagrange density of the Dirac equation is also invariant under this transformation. Now the leptons at this point are massless and will only become massive under spontaneous symmetry breaking. As a result the Dirac Lagrange density is obtained from the usual density by setting the mass equal to zero. Using the left and right-handed split of the wave function above, the resulting Lagrange density,

$$\mathcal{L} = i\overline{\begin{pmatrix}\psi_{\nu l}\\ \psi_l\end{pmatrix}}_L \gamma^\mu\partial_\mu\begin{pmatrix}\psi_{\nu l}\\ \psi_l\end{pmatrix}_L + i\overline{(\psi_l)}_R\gamma^\mu\partial_\mu(\psi_l)_R,$$

will only be gauge invariant if the derivative ∂_μ is replaced by

$$D_\mu = \partial_\mu - ig\boldsymbol{T}\bullet\boldsymbol{W}_\mu - i\frac{g'}{2}YB_\mu.$$

The photon field constructed from the two neutral fields must be combined so that the physical state given by A_μ is massless, and will have the form

$$A_\mu = B_\mu\cos\theta_w + W_\mu^3\sin\theta_w,$$

and the combination orthogonal[†] to A_μ is the combination corresponding to the neutral intermediate boson of the weak interactions

$$Z_\mu = -B_\mu \sin\theta_w + W_\mu^3 \cos\theta_w.$$

The last two equations may be inverted to give

$$B_\mu = A_\mu \cos\theta_w - Z_\mu \sin\theta_w$$
$$W_\mu^3 = A_\mu \sin\theta_w + Z_\mu \cos\theta_w.$$

The mixing angle θ_W is known as the Weinberg angle. Note that Z_μ is often written as Z_μ^0, Z^0, or W^0 in the literature.

The electroweak neutral current interaction, as indicated above, can be written as

$$-igJ_\mu^3 W_\mu^3 - i\frac{g'}{2} j_\mu^Y B_\mu.$$

With a little algebra this can be put into the form

$$-igJ_\mu^3 W_\mu^3 - i\frac{g'}{2} j_\mu^Y B_\mu = -i\left(gJ_\mu^3 \sin\theta_w + g' \frac{j_\mu^Y}{2} \cos\theta_w \right) A_\mu - i\left(gJ_\mu^3 \cos\theta_w - g' \frac{j_\mu^Y}{2} \cos\theta_w \right) Z_\mu.$$

The first bracketed term on the right hand side is the electromagnetic interaction. From the definitions of j_μ^{em}, j_μ^Y, J_μ^3, and Q above, one finds the important relation

$$e j_\mu^{em} = e\left(J_\mu^3 + \frac{1}{2} j_\mu^Y \right).$$

This must be the same as the first term bracket, so that both $g\sin\theta_w$ and $g'\cos\theta_w$ equal e, or $\tan\theta_w = \frac{g'}{g}$. This tells us that the couplings g and g' may be replaced with e and θ_W, where θ_W is determined by experiment.

Again, with a little algebra, the weak neutral current of the second bracket may be written as

---

[†] Draw a set of *x,y*-axes in the plane; label the *x*-axis Z_μ and the *y*-axis A_μ; draw the vector W_μ^3 in the first quadrant at an angle θ with respect to the Z_μ-axis and vector B_μ orthogonal to in the 2nd-quadrant. The expressions for A_μ and Z_μ follow from projecting W_μ^3 and B_μ on the Z_μ and A_μ axes.

$$-i\frac{g}{\cos\theta}\left(J_\mu^3 - j_\mu^{em}\sin^2\theta\right)Z_\mu.$$

The expression in the brackets is defined as the observed weak neutral current J_μ^{NC}. Thus, it is possible to write the observed neutral current J_μ^{NC} as a sum of a left-handed component J_μ^3 of SU(2)$_L$ and a right-handed component taken from the electromagnetic current j_μ^{em}. The electromagnetic current j_μ^{em} in turn may be written as the sum of J_μ^3 contained in SU(2)$_L$ and the weak hypercharge current j_μ^Y, which is invariant under SU(2)$_L$ and has only a right-handed component. The net result is then

$$j_\mu^{em} = J_\mu^3 + \frac{1}{2}j_\mu^Y$$
$$J_\mu^{NC} = J_\mu^3 - j_\mu^{em}\sin^2\theta.$$

Consider now the electron and its neutrino so that the wave functions are

$$\chi_L = \overline{\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}}_L \qquad \text{with } T = \frac{1}{2} \text{ and } Y = -1$$
$$\psi_R = e_R^- \qquad \text{with } T = 0 \text{ and } Y = -2.$$

The electroweak Lagrangian may then be written as

$$\mathcal{L}_1 = \overline{\chi_L}\gamma^\mu\left[i\partial_\mu - \frac{1}{2}g\boldsymbol{\tau}\bullet\boldsymbol{W}_\mu + \frac{1}{2}g'B_\mu\right]\chi_L + \overline{e_R}\gamma^\mu\left[i\partial_\mu + g'B_\mu\right]e_R^- - \frac{1}{4}\boldsymbol{W}_{\mu\nu}\bullet\boldsymbol{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}.$$

where the last two terms correspond to the kinetic energy and self coupling of the fields. This Lagrangian describes massless gauge bosons and fermions. Gauge invariant masses are introduced by use of spontaneous symmetry breaking.

Spontaneous Symmetry Breaking

Spontaneous symmetry breaking is in essence a consequence of redefining the concept and nature of the vacuum. The Standard Model introduces a vacuum significantly different from that of standard quantum field theory. There, the vacuum state, Ψ_0, is the quantum state where no particles are present. It is invariant under a unitary transformation so that $U\,\Psi_0 = \Psi_0$. It was soon recognized, however, that there was a problem with the quantum theory of fields. An operator, such as the electric field $\hat{E}$, is not well defined, whereas "smeared" fields such as

$$\int \hat{E}(\boldsymbol{x}, t)f(x)d\boldsymbol{x}dt = \hat{E}(f),$$

where f is a smooth infinitely differentiable function, are well defined. One possible way of interpreting this difficulty is that it is due to the nature of the vacuum, which will be further discussed later in this essay. The implication is that the quantum field theoretic view of the vacuum must be modified at small distances. This is precisely what lies at the heart of symmetry breaking. By redefining the vacuum local gauge symmetries may be "broken" resulting in gauge bosons gaining mass through the Higgs mechanism.

With some reasonable assumptions on the transformation properties of the Lagrangian, the Noether theorem tells us that if the Lagrangian has some number of symmetries there will be the same number of conserved currents and consequently, because of the equation of continuity, the same number of conserved charges. This fact is the basis for proving an important theorem by Goldstone, which has both a classical and quantum mechanical formulation. The latter can be expressed as follows: If there exists a field operator $\phi(x)$ such that the vacuum expectation value $<0|\ \phi(x)\ |0> \neq 0$, and which does not transform as a singlet (a 1-dimensional representation or a spin-zero state) under some transformation group, then there exist massless particles in the spectrum of states.

The vacuum state, Ψ_0, will no longer be defined as the quantum state where no particles are present. Rather, it is assumed to be analogous to the ground state of an interacting many body system, and will be defined as the state of minimum energy so that the vacuum expectation value of the Hamiltonian $<0|H|0>$ is a minimum. It is the minimum of the potential energy that will play the major role in what follows. Furthermore, in quantum mechanics the ground state is non-degenerate. In the case of quantum field theory, this will no longer be the case. Degenerate orthogonal ground states, where tunneling between them is not possible, will be allowed.

If the Lagrangian is invariant under a gauge group G and after symmetry breaking the vacuum remains invariant under $H \subset G$, that is, a subgroup of G, the number of massless Goldstone bosons is equal to the number of symmetries that are broken. Equivalently, this is equal to the dimension of the coset space $\dim(G/H)$ or the number of generators of G that are *not* generators of H. The magic of symmetry breaking in the Standard Model is that when the symmetry is local, so that the gauge transformations depend on space-time, no Goldstone bosons appear, and instead the symmetry breaking results in massive gauge bosons whose number is equal to dim(G/H). The total number of gauge particles, both massive and massless, is given by dim$G$.

While the Lagrangian of a system may be invariant under some symmetry group, the vacuum state may not be invariant. The classical example of this is the ferromagnet. Above the critical Curie temperature, the spins are randomly oriented and the ground state is spherically symmetric. Below the Curie temperature, the spins for each magnetic domain are aligned and the ground state is no longer spherically symmetric with regard to rotations. The rotational symmetry of the Hamiltonian is "spontaneously broken" to the cylindrical symmetry of each magnetic domain along the direction of its magnetization. The directions are random and have the same energy in the absence of an external magnetic field. The vacuum or ground state is degenerate and does not share the symmetry of the Hamiltonian.

An analogy often used in the literature is that of a superconductor, and in the case of quantum chromodynamics one speaks of "color superconductivity". For the usual superconductor, one can show that combining the London equation for the current with Maxwell's equations leads to a relation for the magnetic field penetration into the surface of a superconductor whose solution is a decreasing exponential. This is known as the Meissner effect. Its importance for spontaneous symmetry breaking is that it transforms the long-range electromagnetic field into one that, in the superconductor, has short range without violating the gauge invariance of Maxwell's equations. In addition, we know that Yukawa showed that short-range forces correspond to massive quanta. The superconductor is said to give the massless photon a "mass" within the superconductor. In this way, superconductivity can be used as an example of spontaneous symmetry breaking in the Abelian Higgs model.

The semi-classical approach here to explaining spontaneously broken symmetries will be to start with the Lagrangian for the field, put in a special form for the potential that redefines the vacuum, and impose symmetry requirements. Recall that if one puts the Lagrange density (often simply called the Lagrangian) into the Euler-Lagrange equation, one obtains the equation of motion. So, for example, substituting the Lagrange density $\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}m^2\phi^2$ into the Euler-Lagrange equation results in the Klein-Gordon equation $\partial_\mu\partial^\mu\phi + m^2\phi = 0$.

Let us begin with *global* symmetry breaking, which means that gauge transformations are not space-time dependent. The simplest example is that of U(1). The general Lagrange density for a complex scalar field $\phi = (\phi_1 + i\phi_2)$ is

$$\mathcal{L} = (\partial_\mu\phi)(\partial^\mu\phi^*) - V(\phi\phi^*).$$

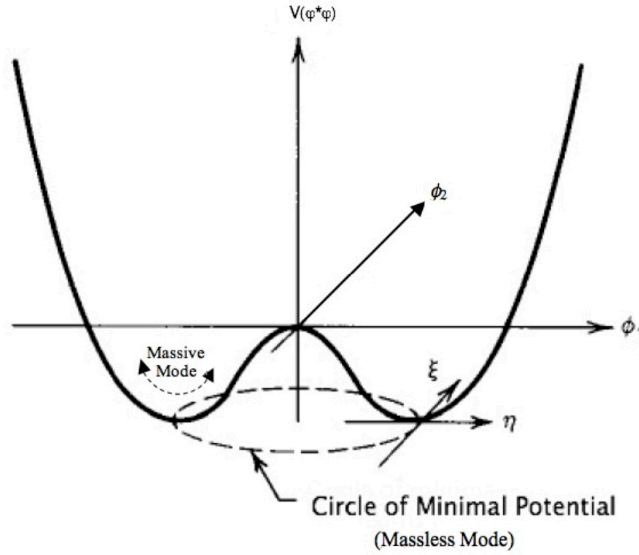
For the potential V, one chooses a form originally proposed by Ginzburg and Landau before the BCS theory of superconductivity. This type of potential was intended to represent the Helmholtz free energy of a second order phase transition. The reason for choosing it here is that this form of potential works to give the desired result (and possibly tells us something about the nature of the vacuum) even though it was intended as a phenomenological description of the free energy density of a superconductor. In gauge theory it provides a type of self-interaction of the Higgs field. As already noted, it also drastically redefines the nature of the vacuum. It is given by

$$V(\phi, \phi^*) = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2.$$

The self-interaction comes from the λ term. The extrema of this function are found by taking the first and second derivative with respect to $|\phi|$ and setting the result equal to zero. Doing the algebra (and using the definition of ϕ) results in

$$-\frac{\mu^2}{2\lambda} = \phi^*\phi = \phi_1^2 + \phi_2^2 =: a^2.$$

a^2 is real for the choice $\lambda > 0$, $\mu^2 < 0$, which we make here. There is also the solution $\phi = \phi^* = 0$. Examining the second derivative tells us that this solution is a relative maximum and that the solution at $a^2 = -\mu^2/2\lambda$ is a relative minimum. A sketch of the potential is shown below



Although the components of ϕ are drawn as coordinates, it should be remembered that ϕ is a field. The minima of the potential lie along the circle of minimal potential of radius a that comprise a set of degenerate vacua related by a rotation about the axis corresponding to the magnitude of the potential. The potential along the circle, in the ξ direction tangent to the circle, is constant. It therefore takes no energy to move along this path and motion along it corresponds to the massless mode, while motion in a plane containing the V-axis does take energy and corresponds to the massive mode.

Let us now transform to polar coordinates so that

$$\phi(x) = \rho(x)e^{i\theta(x)},$$

where x is the space-time coordinate so that the same form holds at each space time point. The vacuum is then $<0|\phi|0> = <0|\rho|0> = a$ and $<0|\theta|0> = 0$. The degenerate vacua are then connected by a U(1) symmetry transformation. Note that the U(1) phase symmetry is destroyed as a result of the vacuum being given by the choice of $\rho = a$ and some particular value of θ; it is the specification of θ that breaks the symmetry. We will be interested in small oscillations around the vacuum state located at the circle of minimal potential. The quanta of these oscillations correspond to physically interesting particles. Because the minimum of the potential lies at a radial distance a from the origin, the following transformation is made:

$$\phi(x) = \left(\rho'(x) + a\right)e^{i\theta(x)}.$$

As a result the vacuum is now <0|ρ' |0> = <0|θ|0> = 0. This ϕ is then substituted into the Lagrangian given above there results, after a bit of algebra, a kinetic term and the potential term

$$V = \lambda\left(\rho'^4 + 4a\rho'^3 + 4a^2\rho'^2 - a^4\right).$$

The quadratic term in ρ' implies that ρ' has a mass of $4\lambda a^2$. Spontaneous symmetry breaking has generated this mass. Notice that there is no similar term in θ^2, implying that θ is a massless field. This can be thought of as being a consequence of there being no restoring force in the θ-direction. ϕ_1 and ϕ_2 started out as two fields satisfying the Klein-Gordon equation. After symmetry breaking we have a massive field ρ' and a massless field θ. This is an example of the group theoretic requirements set out above. Another example or two might be helpful.

Let the symmetry group be SO(3). A Lagrangian for a Lorentz invariant, massive isovector scalar field is

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi_i\partial^\mu\phi_i - \frac{\mu^2}{2}\phi_i\phi_i - \lambda\left(\phi_i\phi_i\right)^2,$$

where i = 1,2,3. After symmetry breaking, we get a degenerate isospin vacuum state, and must choose one. Once this dirction in isospace is chosen the vacuum is no longer invariant under the three generators of SO(3) but only under rotations about the fixed axis in isospace. We started out with three massive fields and after symmetry breaking have only one, corresponding to the fixed axis in isospace. Two Goldstone bosons appear corresponding to the loss of symmetry about the two other axes. Thus, three massive scalar fields result in one massive scalar field and two massless scalar fields.

If we demand that the Lagrangian above be invariant under a *local* rather than global gauge transformation so that $\phi \to e^{i\Lambda(x)}\phi$, the derivative must be changed to the covariant derivative thereby introducing additional terms into the Lagrangian. This results in the Lagrangian taking the form

$$\mathcal{L} = \frac{1}{2}\left(\mathrm{D}_\mu\phi_i\right)\left(\mathrm{D}^\mu\phi_i\right) - \frac{\mu^2}{2}\phi_i\phi_i - \lambda\left(\phi_i\phi_i\right)^2 - \frac{1}{4}F^i_{\mu\nu}F^{i\mu\nu},$$

where

$$\mathrm{D}_\mu\phi_i = \partial_\mu\phi_i + g\varepsilon_{ijk}A^j_\mu\phi_k, \qquad F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g\varepsilon^{ijk}A^j_\mu A^k_\nu.$$

Here we are starting with three massive scalar fields ϕ_i and three massless vector fields A^i_ν. After symmetry breaking, there is again a degenerate vacuum state and after choosing a direction in isospace, say ϕ_3, there remains only one massive scalar field ϕ_3. But instead of two Goldstone bosons, we find that two of the vector fields A^i_ν have become massive and one remains massless. The vacuum remains invariant only under U(1), the group with one generator, corresponding to the one massless vector field.

The Weinberg-Salam Model and Gauge Field Masses

The introduction of the symmetry breaking Higgs field in the Weinberg-Salam electroweak theory, is perhaps its most important feature. Mass is generated for gauge bosons by their *interaction* with the Higgs field. This mechanism is also used to produce quark masses in QCD and thus redefines the nature of the vacuum down to very small distances. The Higgs field has a totally unknown origin and is simply postulated by analogy to examples such as those discussed above.

The electroweak Lagrangian was given above as

$$\mathcal{L}_1 = \overline{\chi}_L\gamma^\mu\left[i\partial_\mu - \frac{1}{2}g\boldsymbol{\tau}\boldsymbol{\cdot}\boldsymbol{W}_\mu + \frac{1}{2}g'B_\mu\right]\chi_L + \bar{e}_R\gamma^\mu\left[i\partial_\mu + g'B_\mu\right]e_R - \frac{1}{4}\boldsymbol{W}_{\mu\nu}\boldsymbol{\cdot}\boldsymbol{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}.$$

The aim now is to use the Higgs formalism to give the $W^\pm$ and Z^0 mass while leaving the photon massless. If a scalar field is to be used for spontaneous symmetry breaking an appropriate Lagrangian, having $\mathrm{SU}(2)\times\mathrm{U}(1)$ gauge invariance, for the scalar field must be added to the above Lagrangian for the scalar field. It will look very similar in form to the above except that the kinetic energy terms will be eliminated. The scalar Lagrangian is

$$\mathcal{L}_2 = \left(\partial_\mu\phi + ig\boldsymbol{T}\boldsymbol{\cdot}\boldsymbol{W}_\mu\phi + i\frac{1}{2}g'B_\mu\phi\right)^\dagger\left(\partial_\mu\phi + ig\boldsymbol{T}\boldsymbol{\cdot}\boldsymbol{W}_\mu\phi + i\frac{1}{2}g'B_\mu\phi\right) - \mu^2\phi^\dagger\phi - \lambda\left(\phi^\dagger\phi\right)^2.$$

Here, ϕ represents four real scalar fields ϕ_i that—following Weinberg—are used to form an isospin doublet with weak hypercharge $Y = 1$; that is,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \qquad \begin{array}{l} \phi^+ := \frac{1}{\sqrt{2}}\left(\phi_1 + i\phi_2\right) \\ \phi^0 := \frac{1}{\sqrt{2}}\left(\phi_3 + i\phi_4\right) \end{array}.$$

The potential with $\mu^2 < 0$ and $\lambda > 0$ will cause spontaneous symmetry breaking, as discussed above, leaving a local SU(2) gauge freedom that can transform ϕ_1, ϕ_2, and ϕ_4 away so that

$$\phi = \begin{pmatrix} 0 \\ \phi_3 + a \end{pmatrix}.$$

The vacuum expectation value, ϕ_0, is then

$$\phi_0 = \left\langle 0 \right| \phi \left| 0 \right\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ a \end{pmatrix}.$$

This is known as the Higgs vacuum or ground state and is assumed *ab initio* to be electrically neutral to guarantee that the photon remains massless. It does however carry weak hypercharge and isotopic spin in order to interact with the Z^0, W^+ and W^- fields. It should also be noted that when these fields are massless, they only have spin components parallel and antiparallel to the momentum, but after symmetry breaking they gain an additional transverse spin component.

Remembering the Gell-Mann-Nishijima relation,

$$Q = T^3 + \frac{Y}{2},$$

we see that with the choice of $T = |\,\boldsymbol{T}\,| = 1/2$, $T^3 = -1/2$, and $Y = 1$ the charge operator Q that generates $U(1)_{em}$ will yield zero when operating on ϕ_0. This means that the vacuum will remain invariant under $U(1)_{em}$ local gauge transformations since for any value of $\alpha(x)$,

$$\phi_0 \rightarrow \phi_0' = e^{i\alpha(x)Q}\phi_0 = \phi_0.$$

Because this is the case, the photon remains massless.

The gauge boson masses are obtained by substituting ϕ_0 into the Lagrangian $\mathcal{L}_2$ above. The relevant term in $\mathcal{L}_2$ with $Y = 1$ is

$$\left(ig\boldsymbol{T}\cdot\boldsymbol{W}_\mu\phi + i\tfrac{1}{2}g'B_\mu\phi\right)^\dagger\left(ig\boldsymbol{T}\cdot\boldsymbol{W}_\mu\phi + i\tfrac{1}{2}g'B_\mu\phi\right).$$

Putting in the matrices and remembering that $T^i = \tau^i/2$, where the τ^i are the Pauli matrices, results in

$$\left(\frac{i}{2}\begin{pmatrix} \left(gW_\mu^3 + g'B_\mu\right) & g\left(W_\mu^1 - iW_\mu^2\right) \\ g\left(W_\mu^1 + iW_\mu^2\right) & \left(-gW_\mu^3 + g'B_\mu\right)\end{pmatrix}\phi_0\right)^\dagger\left(\begin{pmatrix} \left(gW_\mu^3 + g'B_\mu\right) & g\left(W_\mu^1 - iW_\mu^2\right) \\ g\left(W_\mu^1 + iW_\mu^2\right) & \left(-gW_\mu^3 + g'B_\mu\right)\end{pmatrix}\phi_0\right).$$

Going through the matrix algebra yields

$$\tfrac{1}{8}a^2g^2\left[\left(W_\mu^1\right)^2 + \left(W_\mu^2\right)^2\right] + \tfrac{1}{8}a^2\left(g'B_\mu - gW_\mu^3\right)^2.$$

The first term can be written as

$$\left(\tfrac{1}{2}ag\right)^2 W_\mu^+ W^{-\mu}.$$

Since we are working in Minkowski space, the μ index can be raised or lowered as needed to improve clarity. Remembering the Lagrange density, for the Klein-Gordon equation, $\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\right)\left(\partial^\mu\phi\right) - \frac{1}{2}m^2\phi^2$, and comparing the mass term with the above, tells us that $M_W^2 = \left(\frac{1}{2}ag\right)^2$, so that $M_W = \frac{1}{2}ag$.

The second term, $\frac{1}{8}a^2\left(g'B_\mu - gW_\mu^3\right)^2$, is transformed by use of the relations

$$B_\mu = A_\mu\cos\theta_W - Z_\mu\sin\theta_W$$
$$W_\mu^3 = A_\mu\sin\theta_W + Z_\mu\cos\theta_W,$$

described earlier in this essay, into

$$\tfrac{1}{8}a^2\left[A_\mu\left(g'\cos\theta_W - g\sin\theta_W\right) - Z_\mu\left(g\cos\theta_W + g'\sin\theta_W\right)\right]^2.$$

If the photon is to remain massless, the term $\left(g'\cos\theta_W - g\sin\theta_W\right)$ must vanish. This will be the case, and the requirement that $\sin^2\theta_W + \cos^2\theta_W = 1$ will be satisfied, if

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \qquad \text{and} \qquad \cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}.$$

Substituting these into the remaining term gives

$$\tfrac{1}{8}a^2 Z_\mu^2 \left(g\cos\theta_W + g'\sin\theta_W\right)^2 = \tfrac{1}{2}a^2 \frac{\left(g^2 + g'^2\right)}{4} Z_\mu^2 = \tfrac{1}{2}M_Z^2 Z_\mu^2.$$

The mass of the neutral Z_μ is then

$$M_Z = a\frac{\left(g^2 + g'^2\right)^{\frac{1}{2}}}{2}.$$

A neutral vector boson has the same form of mass term as that in the Lagrangian of the Klein-Gordon equation, which can be generalized to several fields having a vector or spinor character.

What has happened here is that the electroweak symmetry group has been dynamically broken to yield the electromagnetic group, that is $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y} \xrightarrow{\mathrm{SB}} \mathrm{U}(1)_\mathrm{EM}$, while keeping the photon massless, along with generating mass for the W and Z bosons.

Thus far we have found the masses of the bosons but not the fermions. The problem is that the mass term in the free field Dirac equation destroys gauge invariance under all gauge transformations and in particular under $\mathrm{SU(2)_L}$. As a result it was excluded in the Lagrangian $\mathcal{L}_1$ above, which was for massless fermions and gauge bosons. In somewhat simplified notation, the electron mass term would be $-m_e\left(\overline{e}_L e_R + \overline{e}_R e_L\right)$. Because e_R is a singlet and e_L a member of an isospin doublet, this term cannot be gauge invariant.

The way around this problem is to introduce a coupling between the originally massless fermion fields and the Higgs field. The procedure for giving mass to the lepton and quark fermions is the same as above although there will be no way to predict the strength of the coupling, which becomes a parameter that must be fixed by the observed masses.

With a slight change in notation from that given above, we defined the wavefunctions

$$\chi_L = \overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}}_L, \qquad \psi_R = e_R.$$

Remembering that the Higgs field is

$$\phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix},$$

and that the complex conjugate of ϕ^{+} is ϕ^{-}, the following Lagrangian must be added to $\mathcal{L}_1$ given above:

$$\mathcal{L}_3 = -G_e\left[\left(\overline{\nu_e}, \bar{e}\right)_{\mathrm{L}} \begin{pmatrix} \phi^{+} \\ \phi_0 \end{pmatrix} e_R + \overline{e_R}\left(\phi^{-}, \overline{\phi_0}\right) \overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}}_L\right].$$

One again goes through the process of symmetry breaking using the potential discussed above and substitutes

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ a + h(x) \end{pmatrix}.$$

The Higgs doublet has been reduced to the neutral field $h(x)$ and after again using the gauge freedom to transform ϕ_1, ϕ_2, and ϕ_4 away, the Lagrangian $\mathcal{L}_3$ becomes

$$\mathcal{L}_3 = -\frac{G_e}{\sqrt{2}}a\left(\overline{e_L}e_R + \overline{e_R}e_L\right) - \frac{G_e}{\sqrt{2}}\left(\overline{e_L}e_R + \overline{e_R}e_L\right)h.$$

The electron mass is then given by choosing G_e such that $m_e = \frac{G_e}{\sqrt{2}}a$.

Masses for the quarks that appear in the QCD discussion below are obtained in the same way except that there are some changes in the choice of the Higgs doublet to generate the mass for the upper member of a quark doublet. The masses associated with the up and down quarks, which make up protons and neutrons are quite small and to a good approximation can be set equal to zero. Most of the mass of the proton and neutron come from the relativistic motion of their constituent quarks and the energy in the color gluon fields that govern their interaction.

Particles of the Standard Model & QCD

Let us now list the fundamental particles of the Standard Model:

Gauge group: $SU(3)_C \times SU(2)_L \times U(1)_Y \Rightarrow 8 + 3 + 1$ Gauge bosons.

The 12 gauge bosons are the $W^{\pm}$, Z^0 , the photon γ, and 8 gluons all having spin equal to 1. Three generations of Quarks and Leptons:

$$u \quad d \quad e^- \quad \nu_e$$

$$c \quad s \quad \mu^- \quad \nu_\mu$$

$$t \quad b \quad \tau^- \quad \nu_\tau$$

In addition, the quarks and gluons carry three conserved color charges (r, b, g), and antiquarks carry anticolor $(\bar{r}, \bar{b}, \bar{g})$. The u, c, and t quarks carry a charge in terms of e of +2/3, while the d, s, and b quarks have a charge of −1/3.

And finally, one must introduce the spin-zero Higgs boson for the Higgs field needed to generate masses.

Excluding the Higgs, this is summarized in the following table:

			Name	Mass [GeV]	Charges		
					electric	isospin	colour
Fermions – spin 1/2	Leptons	ν_e	el. neutrino	$< 10^{-8}$	0	+1/2	0
		e	electron	0.000511	−1	−1/2	0
		ν_μ	muon neutrino	< 0.0002	0	+1/2	0
		μ	muon	0.106	−1	−1/2	0
		ν_τ	tau neutrino	< 0.02	0	+1/2	0
		τ	tau	1.777	−1	−1/2	0
	Quarks	u	up	0.003	+2/3	+1/2	3
		d	down	0.006	−1/3	−1/2	3
		c	charm	1.3	+2/3	+1/2	3
		s	strange	0.1	−1/3	−1/2	3
		t	top	174.3	+2/3	+1/2	3
		b	bottom	4.3	−1/3	−1/2	3
Bosons	spin 1	γ	photon	0	0	0	0
		Z^0	Z boson	91.19	0	0	0
		$W^{\pm}$	W boson	80.42	±1	±1	0
		g	gluon	0	0	0	8

Both left-handed and right-handed quarks form triplets under $SU(3)_C$, while left-handed quarks are doublets under SU(2) and right-handed quarks are singlets under SU(2). Leptons do not participate in strong interactions and are therefore color singlets under $SU(3)_C$.

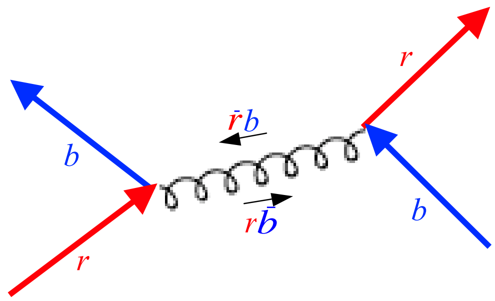
The six leptons of the Standard Model, e, ν_e, μ, ν_μ, τ, ν_τ, interact through electromagnetic and weak forces. The gauge bosons $W^\pm$ and Z^0 associated with $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ are massless, but the weak interactions are known to have a short range, and the form of the Yukawa potential, $e^{-m/r}/r$, tells us that these must be massive. On the other hand, it can be shown that they *must* be massless to preserve gauge invariance. As we have seen, this conundrum is resolved in the Standard Model by spontaneous symmetry breaking with the Higgs field. This results in the $W^\pm$ and the Z^0 gauge bosons acquiring mass, the photon remaining massless, and most importantly, results in a theory that is renormalizable.

All strongly interacting particles are composed of three quarks, while the mesons are a bound state of a quark and antiquark. But then there is a problem since quarks obey Fermi-Dirac statistics so that the Pauli principle forbids the existence of states with three identical quarks. The Δ^{++} is such a state, and to resolve the contradiction the new quantum number of color was introduced.

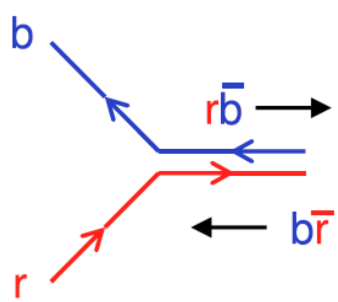
The use of the term color charge can be a bit misleading. Charge in electromagnetism is a scalar quantity. In QCD, color charge is a quantum vector charge and the composite color charge of some configuration of quarks is similar to that of combining angular momentum in quantum mechanics. The three color states form a basis in a 3-dimensional complex vector space. The color state can be rotated by elements of SU(3). The strong interactions in nature rule out states that are not color neutral—all are color singlets.

Quarks interact via gluons, massless, spin one particles that can be either left or right handed, carry color charge, and can therefore interact with each other. Quarks carry a color charge whereas antiquarks carry anticolor. One rule is that there is color conservation at a quark-gluon interaction vertex, which tells us that gluons carry not only a color charge but also an anticolor charge. The colors are often designated as red, blue, and green.

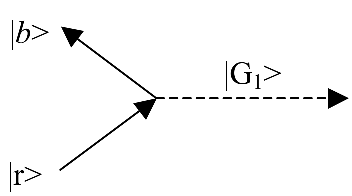
An example of an interaction between two quarks is shown below where only the color exchange is indicated:



The arrows along the gluon symbol indicate from which vertex the gluon originates, so that the $\overrightarrow{r\bar{b}}$ symbol means that one starts from the left vertex and goes to the right one. The opposite is true for the symbol $\bar{r}b$ with the arrow below it. As one can see, there is color conservation at each vertex. Such diagrams for an individual vertex are often drawn as (but, as we will see, the other side of the diagram should not be forgotten):



Let the three color states be represented in Dirac notation as |*r*>, |*b*>, |*g*>. Then the vertex above (changing a red quark to a blue one) going from left to right can be represented as



The Lagrangian will have terms in it that depend on the SU(3) covariant derivative, the key term being $\partial - i\frac{\lambda_k}{2}G_k$, where G_k is the creation operator for a G_k gluon, and the λ_k are the Gell-Mann matrices

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

These matrices are essentially the generators of the SU(3) group.

If one thinks of r, b, g as creation operators that create the states $|r\rangle$, $|b\rangle$, $|g\rangle$ from the vacuum, then the vertex in the figure above has corresponding to it the expression $\langle 0|\,\bar{r}G_1 b\,|0\rangle$, where the terms between the bra and ket vectors should be interpreted, from left to right, as "annihilate an r quark", "create a G_1 gluon", and "create a b quark". The missing vertex on the right hand side will give a similar term $\langle 0|\,\bar{b}G_1 r\,|0\rangle$.

Let the general color state vector in $SU(3)_C$ be represented as $\begin{pmatrix} r \\ b \\ g \end{pmatrix}$, with Hermitian conjugate $\left(\bar{r},\ \bar{b},\ \bar{g}\right)$. Because of the λ_k G_k term, the Lagrangian will contain terms of the form

$$\bar{\Psi}\lambda_k G_k \Psi = \left(\bar{r},\ \bar{b},\ \bar{g}\right)\lambda_k G_k \begin{pmatrix} r \\ b \\ g \end{pmatrix}.$$

This allows us to find the form of the G_k gluons. For example for $k = 1$,

$$\bar{\Psi}\lambda_1 G_1 \Psi = \left(\bar{r},\ \bar{b},\ \bar{g}\right)\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} G_1 \begin{pmatrix} r \\ b \\ g \end{pmatrix} = \bar{b}G_1 r + \bar{r}G_1 b.$$

Consider the first term on the right hand side operating on the vacuum so that we have $\langle 0|\,\bar{b}G_1 r\,|0\rangle$. Similar to the description given for the vacuum expectation value of the second term in the paragraph above, this means, from left to right, "annihilate a b quark", "create a G_1 gluon", and "create a r quark". The vacuum expectation value must be colorless so that $<0|\,\bar{b}G_1 r\,|0>$ implies that $G_1 \sim b\bar{r}$; and $<0|\,\bar{r}G_1 b\,|0>$ implies that $G_1 \sim r\bar{b}$. If G_1 is to fulfill the requirement for both vertices it must then be composed of the terms $r\bar{b}$ and $b\bar{r}$. The combination is generally written $G_1 \sim \frac{1}{\sqrt{2}}\left(r\bar{b} + b\bar{r}\right)$. Going through the same procedure for the rest of the λ_i results in

$$
\begin{array}{lll}
G_1 \sim \frac{1}{\sqrt{2}}\left(r\bar{b} + b\bar{r}\right) & G_2 \sim \frac{i}{\sqrt{2}}\left(b\bar{r} - r\bar{b}\right) & G_3 \sim \frac{1}{\sqrt{2}}\left(r\bar{r} - b\bar{b}\right), \\
G_4 \sim \frac{1}{\sqrt{2}}\left(g\bar{r} + r\bar{g}\right) & G_5 \sim \frac{i}{\sqrt{2}}\left(g\bar{r} - r\bar{g}\right) & G_6 \sim \frac{1}{\sqrt{2}}\left(b\bar{g} + g\bar{b}\right), \\
G_7 \sim \frac{i}{\sqrt{2}}\left(g\bar{b} - b\bar{g}\right) & G_8 \sim \frac{1}{\sqrt{6}}\left(r\bar{r} + b\bar{b} - 2g\bar{g}\right). &
\end{array}
$$

These gluon states are independent in the sense that they cannot be combined to yield one not in the list. Notice that the three by three identity matrix does not appear in the list of λ_i. It would result in a long-range colorless singlet state $\frac{1}{\sqrt{3}}\left(r\bar{r} + b\bar{b} + g\bar{g}\right)$, which does not appear and is not observed in nature. The G_k can also be computed by using

$$
G_k = \left(r,\, b,\, g \right) \lambda_k \begin{pmatrix} \bar{r} \\ \bar{b} \\ \bar{g} \end{pmatrix},
$$

or by defining the "color matrix"

$$
C = \frac{1}{\sqrt{2}} \begin{pmatrix} r\bar{r} & r\bar{b} & r\bar{g} \\ b\bar{r} & b\bar{b} & b\bar{g} \\ g\bar{r} & g\bar{b} & g\bar{g} \end{pmatrix},
$$

and using the expansion $G_k = \mathrm{Tr}\left(\lambda_k^* C\right)$. There are eight gluons because gluons transform in the adjoint representation of SU(3), which is 8-dimensional. Since gluons carry color charge, in principle they can bind together to form colorless states known as "glueballs".

Similar to the $W^{\pm}$ charged bosons in the electroweak case, one may define, for example, $\frac{1}{\sqrt{2}}\left(G_1 \mp iG_2\right) = r\bar{b}$, and similarly for the pairs G_4, G_5 and G_6, G_7. As seen above in the diagrams, these represent the flow of color charge when quarks exchange gluons. There are a set of Feynman rules for QCD that govern quark-quark scattering of various types.

The commuting generators in the λ_i defined above are λ_3 and λ_8. The fact that they commute tells us these generators are associated with two additional simultaneously observable quantum numbers.

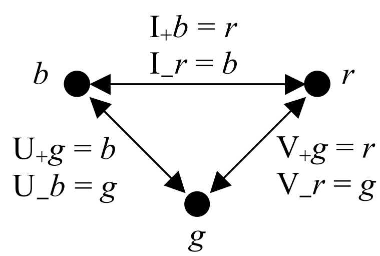
When hadrons are considered, these quantum numbers are defined as $\mathrm{I}_3 = \frac{1}{2}\lambda_3$ and $\mathrm{Y} = \frac{1}{\sqrt{3}}\lambda_8$ giving the weights $\mathrm{I}_3 = \left(+\frac{1}{2}, -\frac{1}{2}, 0\right)$ and $\mathrm{Y} = \left(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}\right)$. Now define the raising and lowering operators

$$\mathrm{I}_\pm = \frac{1}{2}\left(\lambda_1 \pm i\lambda_2\right), \quad \mathrm{U}_\pm = \frac{1}{2}\left(\lambda_6 \pm i\lambda_7\right), \quad \mathrm{V}_\pm = \frac{1}{2}\left(\lambda_4 \pm i\lambda_5\right).$$

Consider the operation of I_+ on the blue color vector:

$$\frac{1}{2}\left(\lambda_1 + i\lambda_2\right)\begin{pmatrix}0\\ b\\ 0\end{pmatrix} = \begin{pmatrix}0&1&0\\0&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}r\\0\\0\end{pmatrix}.$$

The I_+ operator thus converts blue to red. Similarly, using the other raising and lowering operators and color vectors results in the scheme shown below.



The $SU(3)_C$ wavefunctions for the combinations like $r\bar{b}$ are the exact analogues of the SU(3) wavefunctions for quark-antiquark combinations like $d\bar{d}$. With the identification

$$\begin{pmatrix}r\\ b\\ g\end{pmatrix} \Leftrightarrow \begin{pmatrix}u\\ d\\ s\end{pmatrix},$$

one can define color hypercharge and isospin for the three color and three anti-color states for quarks:

	Y_C	I_3^C		Y_C	I_3^C
r	$1/3$	$1/2$	$\bar{r}$	$-1/3$	$-1/2$
b	$1/3$	$-1/2$	$\bar{b}$	$-1/3$	$1/2$
g	$-2/3$	0	$\bar{g}$	$2/3$	0

These color isospin and hypercharge charges should not be confused with flavor isospin and hypercharge for the quarks. *We now shift to a discussion of quarks.*

Because the quark masses compared to that of hadrons are very small, the flavor-independent color force dominates their interactions. Thus if we choose the quark basis as

$$u = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad d = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad s = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

we can, remembering the weights given for λ_3 and λ_8 above, make the following table (compare with the table for color just above):

	Q	Y	I_3		Q	Y	I_3
u	$2/3$	$1/3$	$1/2$	$\bar{u}$	$-2/3$	$-1/3$	$-1/2$
d	$-1/3$	$1/3$	$-1/2$	$\bar{d}$	$1/3$	$-1/3$	$1/2$
s	$-1/3$	$-2/3$	0	$\bar{s}$	$1/3$	$2/3$	0

Q is again the electric charge. Quarks have baryon number 1/3, while antiquarks have baryon number −1/3, so that mesons, composed of a quark and antiquark, have baryon number zero. The way Q is determined is to make use of the same definitions given above in the discussion of color, $I_3 = \frac{1}{2}\lambda_3$ and $Y = \frac{1}{\sqrt{3}}\lambda_8$. In doing so we are using the two additional simultaneously observable quantum numbers also mentioned above. Then, remembering the Gell-Mann-Nishijima relation,

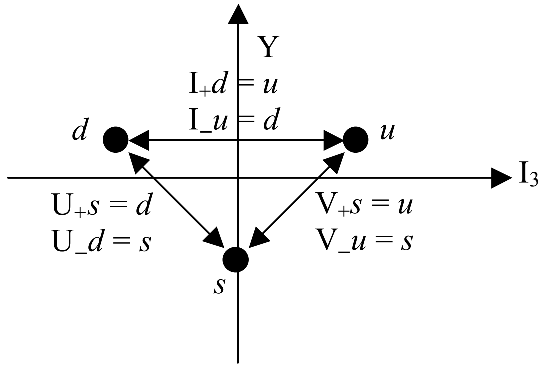
$$Q = I_3 + \frac{Y}{2},$$

we see that Q is given by the matrix

$$Q = \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & -\frac{1}{3} \end{pmatrix}.$$

Given the definition of the quark basis vectors above, we see that the u, d, and s quarks will have electric charges 2/3, -1/3, and -1/3 respectively.

The weight diagram will look similar to that for color above:



The quarks are located at positions consistent with the table above. In any given irreducible representation, the states corresponding to the various particles are characterized by the eigenvalues of I_3 and Y. The effect of the "shift operators" on these states follow from the commutation relations and can be summarized as:

$I_\pm$ results in the changes $\Delta Y = 0$, $\Delta I_3 = \pm 1$, i.e.,

$$I_\pm |I_3, Y\rangle \propto |I_3 \pm 1, Y\rangle.$$

$U_\pm$ results in the changes $\Delta Y = \pm 1$, $\Delta I_3 = -(\pm 1/2)$, i.e.,

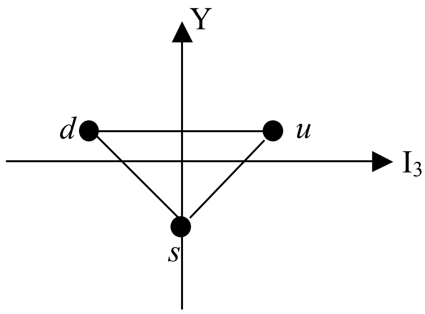
$$U_\pm |I_3, Y\rangle \propto |I_3 - (\pm)1/2, Y \pm 1\rangle.$$

$V_\pm$ results in the changes $\Delta Y = \pm 1$, $\Delta I_3 = \pm 1/2$, i.e.,

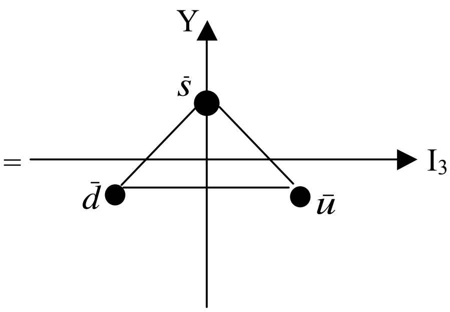
$$V_\pm |I_3, Y\rangle \propto |I_3 \pm 1/2, Y \pm 1\rangle.$$

A simplified version of the diagram above is called the fundamental representation for quarks and is designated as **3,** and the diagram for antiquarks, is designated as $\bar{\mathbf{3}}$, as shown below. Often, the axes and quark labels are also omitted, but are shown in this figure for clarity.

3 =

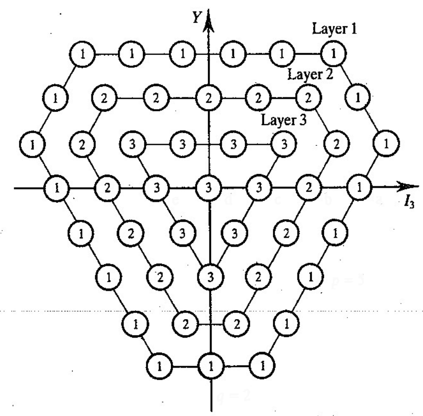


$\bar{\mathbf{3}}$ =



The scale for Y and I_3 in such diagrams is chosen so that the shift resulting from the application of the shift operator V_+ is inclined by 60^o.

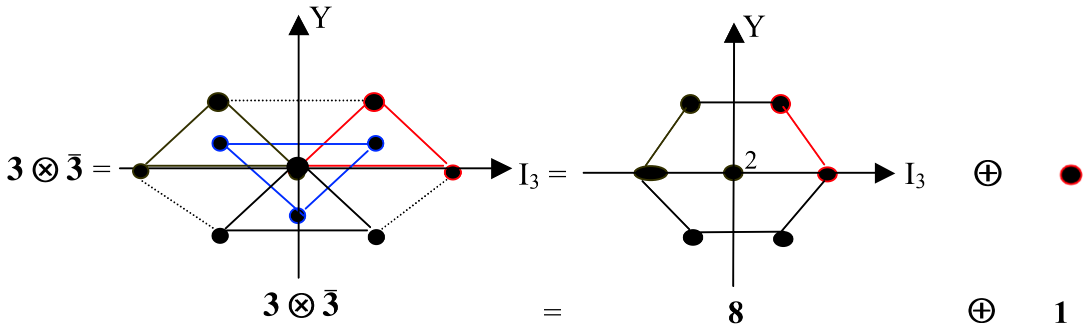
Now we come to a few aspects of what might be termed "diagramatica". To specify a representation or multiplet of SU(3) one must give the sites in the Y-I_3 plane, as is done in the figures above, which are to be occupied, and with what multiplicity (how many states can occupy a given position). This is known as a weight diagram. Generally multiplets are triangular as above or have a hexagonal symmetry as shown below for the large multiplet **81**.



The circles in the figure designate the multiplicity for each site, which are the same on a give layer. The rule is that every site on and inside a boundary are occupied by at least one state, and that when a triangular layer is reached the multiplicity on and inside the triangle have the same multiplicity. Each pattern that satisfies this rule corresponds to only one irreducible representation of SU(3). The fact that multiplicities greater than one occur means that another quantum number in addition to Y and I_3 is needed to distinguish them. The choice made is the total isotopic spin so that any state of an SU(3) irreducible representation is completely and uniquely characterized as $\left| I, I_3, Y \right\rangle$.

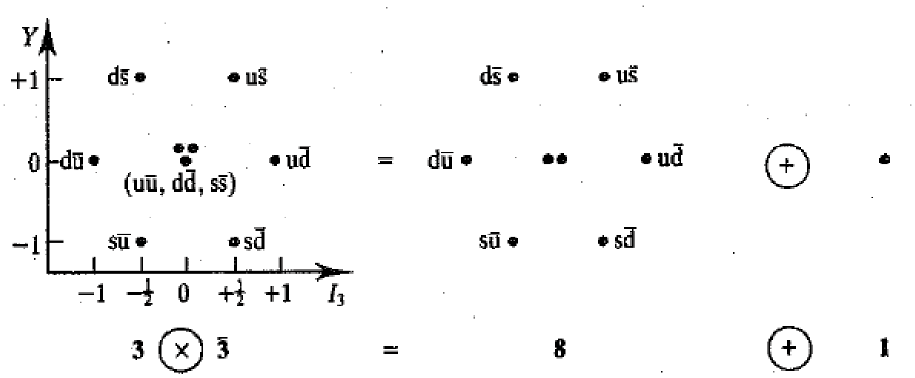
Representation products

Multiparticle states are obtained by taking the product of irreducible representations. There is a graphical method of obtaining products of representations best illustrated by an example. Consider the product $\mathbf{3} \otimes \bar{\mathbf{3}}$. Graphically, the product is interpreted as "put the origin of $\bar{\mathbf{3}}$ on each node of $\mathbf{3}$. The result looks like the figure below.

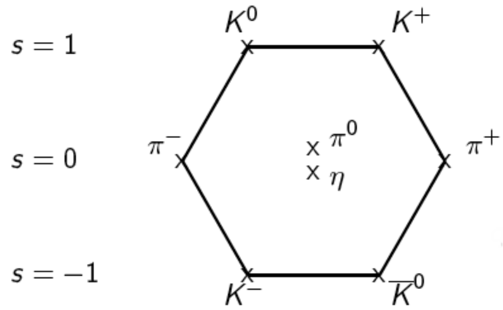


Comparison with the general multiplet figure above, shows that the weights on the boundary (formed by adding the dotted lines) have unit multiplicity, as they should. The original **3** diagram in blue is now eliminated leaving the center with a multiplicity of three. According to the second part of the rule, for an irreducible representation it should be two, so this nonet must reduce to an octet and a singlet as shown in the figure on the right.

The way the various quarks fit into this scheme can be seen by plotting the values for Y and I_3 of the quarks and antiquarks given in the table above. The figure above then becomes

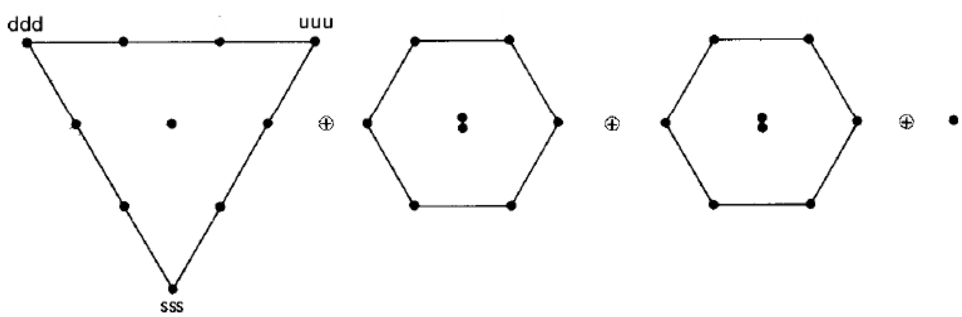


This example can be applied to the meson octet composed of quark-antiquark pairs, as shown below (s is the strangeness, and the I_3 values running from left to right are −1, −1/2, 0, 1/2, and +1.



Each of the mesons corresponds to a colorless quark-antiquark combination.

We have seen above that the properties of SU(3) gave us the result that $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8}$. Similarly, if we want to combine three quarks to form baryons we would have $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10}$. This can be seen in two steps: first $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \bar{\mathbf{3}}$; followed by $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{3} \otimes \left(\mathbf{6} \oplus \bar{\mathbf{3}}\right) = \mathbf{3} \otimes \mathbf{6} \oplus \mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{3} \otimes \mathbf{6} \oplus \mathbf{8} \oplus \mathbf{1}$. One can then show using the graphical techniques above that $\mathbf{3} \otimes \mathbf{6} = \mathbf{8} \oplus \mathbf{1}$, so the result stated follows. It looks like:



$\mathbf{10} \quad \oplus \quad \mathbf{8} \quad \oplus \quad \mathbf{8} \quad \oplus \quad \mathbf{1}$

Note that the center of each figure is at Y = 0, $I_3 = 0$ and I_3 changes in steps of ½ from the origin so that the states indicated by the dots have I_3 values ranging from $-3/2$ to $3/2$. The decuplet figure on the left consists of an I = 3/2 quartet with Y = 1, a triplet with I = 1 and Y = 0, an isospin doublet with Y= -1, and an isospin singlet (I = 0) with Y= -2.

Were SU(3) a perfect symmetry, the 27 particles associated with this product would have the same mass, but the symmetry is not perfect. The symmetry is broken by the large mass of the strange quark compared to that of the up and down quarks.

The completely antisymmetric singlet has a form that can be expressed as $\frac{1}{\sqrt{6}}\left(uds - \mathrm{usd} + sud - sdu + dsu - dus\right)$. There is an analogous color singlet with the form $\frac{1}{\sqrt{6}}\left(rgb - grb + gbr - bgr + brg - rbg\right)$. This color singlet is common to all hadrons and in particular to all baryons. Earlier it was noted that the Δ^{++} with J_3 = 3/2 is described by the symmetric wave function $\mathrm{u}\uparrow\mathrm{u}\uparrow\mathrm{u}\uparrow$. This is symmetric, but it should be antisymmetric under exchange of identical quarks. The addition of the color singlet state to the overall wave function makes it antisymmetric. In general the inclusion of this color singlet in the over all wave function means that only symmetric representations of the remaining product of factors (space ×spin×flavor) in the wave function can be used.

There is a great deal more to the graphical representation used above and its relation to group theory. But the brief introduction given above is perhaps enough to set the stage for approaching the literature and the physics as it evolved historically. In addition many topics have been ignored, such as quark orbital angular momentum and complexities related to spin and quark magnetic moments.

The Standard Model is perhaps the greatest achievement of modern physics. This essay does not even begin to cover the enormous work of the many hundreds, if not thousands, of people over many years that allowed the many experimentally observed particles to fit into the group structure of the Standard Model. Almost all of the physics involved in this achievement has been ignored.

There are, however, some issues that the Standard Model does not even address as well as significant conceptual problems raised by the nature of the vacuum. We begin with the issue of color confinement and then move on to the broader conceptual problems.

Color confinement

Because the gluons of QCD carry color charge, unlike photons, they have three and four gluon self-interactions as illustrated below.

These interactions are thought to be responsible for color confinement, asymptotic freedom (meaning that the quark-quark interaction becomes weaker at short distances allowing perturbation theory to be applicable), and chiral symmetry breaking. Zero-mass quarks would travel at the speed of light and their spin can be aligned either along the direction of motion or opposite to it. This handedness, or chirality, is Lorentz invariant, and this symmetry is explicitly broken when the quark mass is not neglected.

The color potential for quark-quark and quark-antiquark interactions is given by

$$\text{quark–quark: } V(r) = +C\frac{\alpha_S}{r} \qquad \text{quark–antiquark: } -C\frac{\alpha_S}{r}.$$

An overall minus sign, including that of C, is binding. The constant C can be determined from the components of the λ_i matrices, summed over the complete set of matrices, as

$$C(ik \to jl) = \frac{1}{4}\sum_{a=1}^{8} \lambda_{ij}^{a}\lambda_{kl}^{a}.$$

One readily calculates that, for example,

$$\begin{array}{lll} C(rr \to rr) = \frac{1}{3} & C(rg \to rg) = -\frac{1}{6} & C(rg \to gr) = \frac{1}{2} \\ C(r\bar{r} \to r\bar{r}) = \frac{1}{3} & C(r\bar{g} \to r\bar{g}) = -\frac{1}{6} & C(r\bar{r} \to g\bar{g}) = \frac{1}{2}. \end{array}$$

One can show that for color singlet mesons given by the wavefunction

$$|\Psi> = \frac{1}{\sqrt{3}}\left(r\bar{r} + g\bar{g} + b\bar{b}\right),$$

one gets the potential

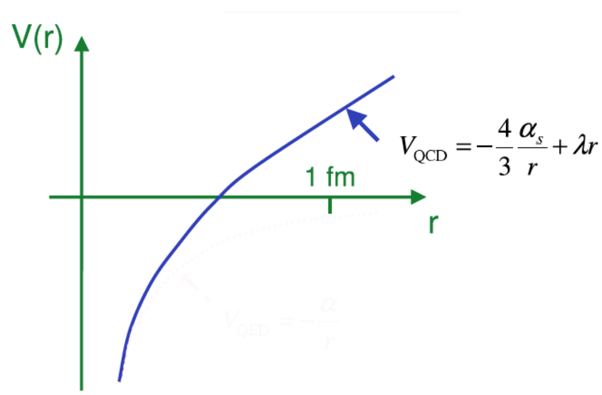
$$V_{q\bar{q}}(r) = -\frac{4}{3}\frac{\alpha_S}{r},$$

which is negative so that for short ranges the color singlet mesons are bound. On the other hand, for the quark-antiquark states in the color octet, for example $|\,\Psi\rangle = |\,r\bar{b}\rangle$ and for which $C(r\bar{b} \to r\bar{b}) = -\frac{1}{6}$, one gets the potential

$$V_{q\bar{q}}(r) = \frac{1}{6}\frac{\alpha_S}{r},$$

so that the *short*-range potential is repulsive.

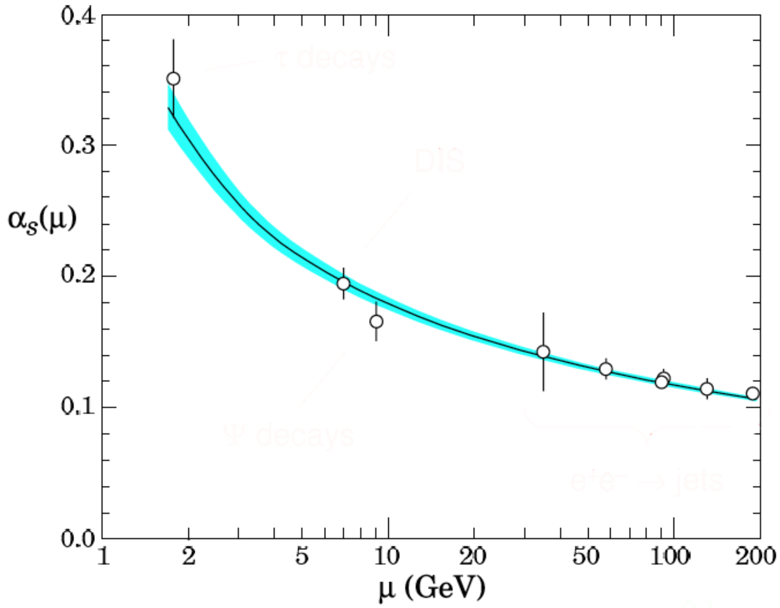
The overall quark-antiquark potential, important for quark and color confinement, is illustrated below:



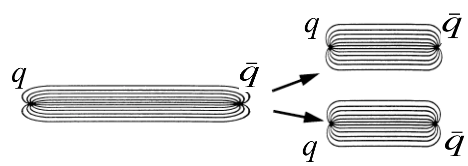
At small distances, $r < 1$fm, the potential is coulomb like, but it becomes proportional to the separation at larger distances, corresponding to $r > 1$fm. λ is the constant of proportionality and α_s is the strong coupling constant, which is not a true constant but rather a "running constant" that decreases with increasing Q, the 4-vector energy momentum transfer of the interaction. To first order it is given by

$$\alpha_s = \frac{12\pi}{(33 - 2N_f)\ln(Q^2/\Lambda^2)},$$

where N_f is the number of allowed quark "flavors" and Λ is an experimentally determined scale parameter (~0.2 GeV). Experimentally, α_s as a function of energy looks like:



The linear nature of the potential at large distances compared to 1fm tells us that the lines of force between gluons are squeezed together into a "flux tube" having constant energy density per unit length governed by the constant λ. The chromoelectric field thus has a string-like character as shown below. As a result, quark-antiquark pairs cannot be separated, but as the distance between them increases, the energy stored in the gluon field will exceed a threshold where the pair will break into two pairs, a process known as hadronisation or fragmentation. A notional idea is given by the following sketch:



Note that the flavor of the quarks in this figure is not specified. If the flavor is the same for the quark and antiquark, it is possible for the pair to mutually annihilate.

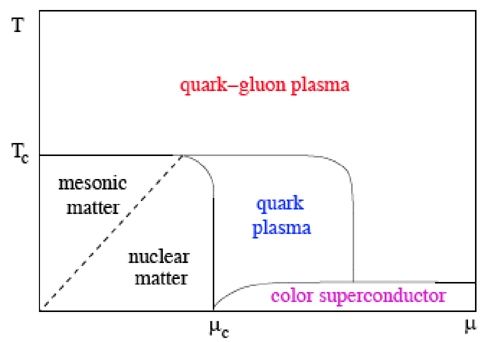
One explanation for the formation of the flux tube is that gluon-gluon self-interactions squeeze the flux lines together. Individual gluons that comprise the "chromoelectric field" illustrated in the figure carry color. Color-anticolor pairs of gluons are color neutral and should strongly attract each other. Also, anti-symmetric states of unlike color charges are attractive while symmetric states are repulsive.

Another possibility is that there exists a vacuum screening current—in analogy with the screening current for magnetic fields in conventional BCS superconductors—except that it would be the chromoelectric field that would be excluded or confined to the equivalent of a single vortex flux line in a type II superconductor. (An array of flux lines would not be expected to be an analogue since the individual flux vortices in a conventional type II superconductor are mutually repulsive.)

Yet another approach to understanding color flux-tube formation is to think of the vacuum as a dual color superconductor. The term comes from the fact that in electromagnetism the duality operation exchanges electric and magnetic fields, and here a dual superconductor is defined as a superconductor in which the roles of the electric and magnetic fields are interchanged. In the usual superconductor, it is the condensation of Cooper pairs formed from the pairing of negatively charged electrons that results in the superconductivity responsible for the Meissner effect; the concept of a dual superconductor has magnetic charges instead of electrons that form boson pairs that condense to form a color superconductor that is thought to be responsible for the analogue of the BCS Meissner effect, and would therefore be expected to exclude color electric fields. It is Gauss' law that prevents the color-electric field from disappearing completely. The most important difficulty with this idea is that there is no evidence that color-magnetic charges exist, and the contention that their condensation would lead to the confinement of quarks remains highly speculative. Even so, the vacuum does seem to have the properties, at least conceptually, of such a dual superconductor.

There is some expectation that a color superconductor, albeit not a dual color superconductor, could exist. It is possible that matter *at ultra-high densities* could be described in terms of a color superconductor. In the central regions of some stars, baryons could approach close enough that their wave functions would overlap, and with increasing density the quarks comprising the baryons would become mobile, ultimately resulting in a quark-gluon plasma, for which there is experimental evidence. The explanation of superconductivity goes something like this: At sufficiently low temperature and high density the quarks would form a degenerate Fermi liquid. Because quarks interact only weakly at short distances (asymptotic freedom), quarks near the Fermi surface are almost free and, unlike electrons in a BCS superconductor, already have a weak attraction. It is argued that the weak attraction under these conditions is sufficient to allow quarks to pair up as bosons so that they could undergo condensation for color superconductivity to appear. Because pairs of quarks cannot be color neutral, the condensate will break the local color symmetry making the gluons massive. The quark pairs would play the same role as the Higgs particle in the standard model.

Here is a notional idea of what the phase diagram for such strongly interacting matter might look like:



T_c is the critical temperature and μ the chemical potential, defined as the partial derivative of the energy with respect to particle number at constant entropy and volume. As one can see, for low enough temperature and high enough density, a color superconductor is expected to form.

Problems with the Standard Model

There are a number of questions that are not answered in the Standard Model: Why are there three families of quarks and leptons? What is the relationship, if any, between quarks and leptons? There are three arbitrary coupling constants associated with the constituent gauge groups of the Standard Model whose value has to be put in by hand. Because the Weinberg mixing angle is arbitrary, there is significant mixing—making the weak and electromagnetic forces appear related—only because experiment shows the coupling constants are of the same order of magnitude. The situation would be different if the mixing angle was close to zero or $\pi/2$. The quantization of charge is not explained since it is put into the theory arbitrarily when assigning values to the weak hypercharge. The Standard Model requires only one Higgs boson, but going beyond the model there may be an expanded "Higgs sector" with a number of Higgs bosons, neutral as well as charged. At this point there is no *strong* evidence for an expanded Higgs sector. In the Standard Model, neutrino masses are zero; yet there is good experimental evidence for a small neutrino masses and for neutrino oscillations—where neutrinos change their flavor. The most popular approach to these problems is to assume the fields of the Standard Model are fundamental, but that they are related by additional symmetries that are broken at higher energy scales. None have yet proved satisfactory.

A few metaphysical thoughts

Perhaps the greatest fundamental problem with the Standard Model is that its redefinition of the vacuum begins to make it look like some form of aether, albeit a relativistic one! One begins to wonder whether the imposition of analogies from condensed matter physics, and in particular superconductivity, are not unwarranted. Surely they should not be taken literally. The fact that they "work", in the sense of supplying an intuitive mechanism, should instead be simply taken as a hint about the real nature of the vacuum. The "vacuum", of course, is just another name for the space-time continuum in the context of quantum field theory, and about which we know very little except for what hints we have from relativity and those given by its definition in the Standard Model.

That there were some problems with the foundations of quantum field theory were mentioned earlier in this essay. There are in fact, far deeper problems than were discussed above, which derive from the well-known fact that the basic assumptions of QFT are inconsistent. The essence of the problem is Haag's theorem, which raises serious questions about the interaction picture that forms the basis for perturbation theory. Because there is a direct bearing on the vacuum, it is worth going into the problem, at least to some extent.

With some simplifications, and simply to emphasize their reasonableness, the usual postulates of QFT are:

1. The state vectors of the quantum system form a separable, normalizable Hilbert space with positive definite metric. State vectors are related to each other by unitary representations of the system's symmetries.
2. The Hilbert space has a vacuum state |0>, which is invariant under Poincaré and any other symmetry transformations associated with the system.
3. The fields $\varphi(x)$ are "smeared" in the sense that there exists an operator $\varphi(f) = \int f(x)\varphi(x)\,dx$, whose domain in Hilbert space is a linear manifold containing the vacuum state. Under a Poincaré transformation, this linear manifold is mapped onto itself, and the smeared fields transform covariantly.
4. The field operators at space-like separated points either commute or anticommute with each other. This is essentially a locality postulate.
5. The result of applying all polynomials of the smeared fields onto the vacuum state |0> results in a dense set in the linear manifold of postulate 3.

If, in addition, we require (a) that the equal time commutation relations are true for the fields, (b) that these commutation relations do not permit inequivalent representations, and (c) that asymptotic fields are in the Hilbert space, then Haag's theorem states that the resulting field theory is for non-interacting particles. Or put another way, Haag's theorem states that the interaction picture exists only if there is no interaction. There are various proofs of Haag's theorem, but except for philosophers of science, the theorem has generally been ignored after the 1970s.

The way chosen by most theorists out of this conundrum is to allow inequivalent representations; i.e., give up (b). One then attempts to use the dynamics of the system to choose one representation from all possible inequivalent representations. That the different representations are unitarily inequivalent means that there is no longer, for the theories associated with each representation, an isomorphism between the states or the observables of the two theories.

When dealing with inequivalent representations, the assumption of the uniqueness of the vacuum is only valid in one particular representation. As a result, giving up (b) leads to a degeneracy of the vacuum. This results in a theory with a broken symmetry. It is extremely interesting, especially from the Standard Model point of view, that symmetry breaking already occurs when choosing the most acceptable way to deal with Haag's theorem. We now give a simple example of a symmetry that does not have a unitary implementation.

Consider the Lagrangian $\frac{1}{2}\partial_\mu\phi\partial^\mu\phi$, which is invariant under the global translation $\phi(x) \to \phi(x) + \chi$, where χ is a space-time independent constant. This invariance applies to the Lagrangian but it cannot be unitarily implemented in the space of states. To show this we will assume that such a unitary transformation exists and show that this assumption results in a contradiction. First of all, using the standard normal mode expansion, we know that $<0|\phi(x)|0> = 0$. If a unitary operator did exist for the transformation $\phi(x) \to \phi(x) + \chi$ one could write

$$\phi'(x) = \phi(x) + \chi = U(\chi)\phi(x)U^\dagger(\chi),$$

where $U = \exp(i\chi Q)$ and Q is a Hermitian generator of the transformation. Q operating on the vacuum state would be expected to result in the eigenvalue 0, so that $<0|U(\chi) = <0|$ and $U^\dagger(\chi)\,|0> = |0>$ as well (expand the exponential). Since $<0|\phi(x)|0> = 0$, the vacuum expectation value of $\phi'(x)$ would be $<0|\, \phi'(x)|0> = \chi$. But from the above equation,

$$<0|\, \phi'(x)|0> = <0|U(\chi)\, \phi(x)\, U^\dagger(\chi)|0> = <0|\phi(x)|0> = 0,$$

so there are two values for the vacuum expectation of $\phi'(x)$, a contradiction that implies that a unitary operator does not exist. Compare this with the discussion of spontaneous symmetry breaking in the earlier part of this essay. We see that when the Lagrangian has a symmetry that cannot be represented in terms of unitary operators on the state space the symmetry is hidden or spontaneously broken.

The electro-weak and QCD symmetry breakings of the Standard Model are generally interpreted as phase transitions in the early, expanding universe that took place after what Fred Hoyle derogatively called "the big bang". As the universe cooled, it presumably passed through some critical temperatures corresponding to the energy scales of these transitions. The various parameters at different times are expected to look like:

	Time t (s)	Energy $E = kT$ (GeV)	Temp. T (K)	Diam. of universe, R (cm)
Planck time, t_{Pl}	10^{-44}	10^{19}	10^{32}	10^{-3}
$SU(2)_L \otimes U(1)$ breaking, M_W	10^{-10}	10^{2}	10^{15}	10^{14}
Quark confinement, $p\bar{p}$ annihilation	10^{-6}	1	10^{13}	10^{16}
ν decoupling, e^+e^- annihilation	1	10^{-3}	10^{10}	10^{19}
Lighter nuclei form	10^{2}	10^{-4}	10^{9}	10^{20}
γ decoupling, transition from radiation cosmos to matter cosmos, atomic nuclei form, stars and galaxies form	10^{12} ($\approx 10^5$ y)	10^{-9}	10^{4}	10^{25}
Today, t_0	$\approx 5 \cdot 10^{17}$ ($\approx 2 \cdot 10^{10}$ y)	$3 \cdot 10^{-13}$	3	10^{28}

It was explained above that zero-mass quarks would travel at the speed of light and their spin could be aligned either along the direction of motion or opposite to it. This handedness, or chirality, is Lorentz invariant, and this symmetry is explicitly broken when the quark mass is not neglected. In the cooling universe scenario, the critical temperature T_c , corresponding to the chiral and confinement transitions, when quarks become bound, are thought to be similar. At $T > T_c$, chiral symmetry is obeyed, and the vacuum expectation value of the quark-antiquark condensate, $\langle 0 \mid q\bar{q} \mid 0 \rangle$, is zero. As the temperature falls below T_c , chiral symmetry is spontaneously broken, and the vacuum expectation value of the condensate becomes nonzero.

The result of a non-zero vacuum expectation value is that the vacuum energy density, associated with such condensates, is enormous. Some estimates from the literature are: $\varepsilon_{vac}^{EW} \sim 10^{46}$ erg/cm^3; $\varepsilon_{vac}^{QCD} \sim 10^{36}$ erg/cm^3; provided we set $V(\phi) = 0$ for $\phi = 0$, and take the Higgs coupling constant as roughly the square of the fine structure constant, the Higgs vacuum energy density would be $\varepsilon_{vac}^{Higgs} = -10^{43} \mathrm{erg/cm}^3$.

If the electroweak and QCD symmetry breakings of the Standard Model are taken to be phase transitions in the early, expanding universe, then general relativity must apply. If, further, the vacuum energy densities are real, they must appear in Einstein's equations in conjunction with the consmological constant. For the *static* Einstein universe, the relation between the radius of curvature of the universe and the cosmological constant is

$$8\pi G\rho = \frac{1}{a^2} = \Lambda,$$

where ρ is the mass density of the dust filled universe (with zero pressure) and a is the radius of curvature. For the vacuum energy densities associated with the electroweak sector or QCD, this equation tells us that the universe would essentially shrivel to almost nothing, or as famously attributed to Pauli, the radius of the world "nicht einmal bis zum Mond reichen würde" [would not even reach to the moon].

Turning to the modern context, we would write an equation relating the effective cosmological constant to the various vacuum energy densities as follows:

$$\Lambda_{eff} = \Lambda_0 + \frac{8\pi G}{c^4}\langle \varepsilon_{vac} \rangle,$$

where Λ_0 is Einstein's original cosmological constant, and ε_{vac} includes contributions from any zero-point energies, vacuum fluctuations, the Higgs field, and QCD gluon and quark condensates.

The experimental value for the energy density associated with the cosmological constant is $\varepsilon_{vac}^{CosCnst} = 10^{-8}\,\text{erg/cm}^3$. The conclusion that must be drawn from this is that the equation above for Λ_{eff} is wrong; that is, in terms of gravitation, the various vacuum energy contributions are effectively zero, either because they represent artifacts of the quantum theories or because they are cancelled out by some unknown mechanism. It would be extremely unlikely that the negative vacuum energy associated with the Higgs, with its arbitrarily chosen zero ($V(\phi) = 0$ for $\phi = 0$), along with similar negative contributions, would exactly cancel out the remaining positive energy contributions.

The evidence given to support the reality of the various contributions to the vacuum energy is the Casimir effect, which is a consequence of the lowest order vacuum fluctuations, and higher order effects like the Lamb shift. But there are alternative explanations. The Casimir effect could result from fluctuations associated with the constituents of the plates rather than vacuum fluctuations. Schwinger's source theory takes this point of view and avoids vacuum fluctuations in both the Casimir and higher order QED effects. As Schwinger put it, ". . . the vacuum is not only the state of minimum energy, it is the state of zero energy, zero momentum, zero angular momentum, zero charge, zero whatever." Pauli also seemed to agree with this position when commenting on field fluctuations in quantum field theory, ". . . it is quite impossible to decide whether the field fluctuations are already present in empty space or only created by the test bodies."

The possibility has also been raised that one should allow the vacuum to have a negative energy spectrum as is done in the Dirac hole theory (of course, Dirac filled up these states); the idea being that positive vacuum energy density contributions would be exactly cancelled by compensating negative energy contributions. Interestingly enough, this possibility is not seriously considered in the literature despite the fact that Schwinger long ago showed that if QFT is to be gauge invariant there is a term (called the Schwinger term) that must vanish. He then showed that if the term does vanish, the vacuum state couldn't be the state with lowest field energy. As put by Schwinger, ". . . it is customary to assert that the electric charge density of a Dirac field commutes with the current density at equal times, since the current vector is a gauge-invariant bilinear combination of the Dirac fields. It follows from the conservation of charge that the charge density and its time derivative, referring to any pair of spatial points at a common time, are commutative. But this is impossible if a lowest energy state—the vacuum—is to exist." Since the argument is rather opaque in the 1959 *Physical Review Letters* article, a derivation is given in Appendix II below for those readers interested in the details.

It is not my intent to explore any particular solution to this "cosmological problem", but rather simply point out that it is a serious issue lying at the foundation of the various field theories that has not really begun to be resolved. Both quantum field theory and general relativity have had spectacular success in explaining the domains of their applicability, but there is no experimental evidence that there is a close relationship between the two. All such claims that a relationship exists are based on theoretical expectations.

Appendix I: Spinor Representations of the Lorentz Group

If $\Lambda^{\mu}{}_{\nu}$ is a Lorentz transformation one has

$$\eta_{\alpha\beta}\Lambda^{\alpha}{}_{\mu}\Lambda^{\beta}{}_{\nu}x^{\mu}x^{\nu}.$$

Because this is true for all x^{μ},

$$\eta_{\mu\nu} = \Lambda^{\alpha}{}_{\mu}\eta_{\alpha\beta}\Lambda^{\beta}{}_{\nu}.$$

The group of matrices satisfying this relation is a Lie group called O(3,1). The elements of O(3,1) that can be built up infinitesimally from the identity is a subgroup called SO(3,1). Thus, the Lorentz transformations infinitesimally close to the identity must have the form

$$\Lambda^{\mu}{}_{\nu} = \delta^{\mu}_{\nu} + \omega^{\mu}{}_{\nu},$$

where $\omega^{\mu}{}_{\nu}$ is a matrix of infinitesimal coefficients. Inserting this into the previous equation for $\eta_{\mu\nu}$ shows that $\omega_{\mu\nu}$ is antisymmetric on its indices. With the convention that $\eta^{00} = -1$ and $\eta^{ii} = 1$, the most general form for $\omega^{\mu}{}_{\nu}$ is

$$\omega^{\mu}{}_{\nu} = \begin{pmatrix} 0 & b_1 & b_2 & b_3 \\ b_1 & 0 & -r_3 & r_2 \\ b_2 & r_3 & 0 & -r_1 \\ b_3 & -r_2 & r_1 & 0 \end{pmatrix}.$$

The b's give infinitesimal boosts in the subscripted directions and the r's rotations about the indicated axes.

To operate on Hilbert space, each element $\Lambda \in SO(3,1)$ must have associated with it a unitary operator U satisfying

$$U(\Lambda_1)U(\Lambda_2) = U(\Lambda_1\Lambda_2).$$

Infinitesimally close to the identity, these operators can be expanded as

$$U(\omega) = I + \frac{i}{2}\omega_{\mu\nu}J^{\mu\nu} + O(\omega^2),$$

where the operators $J^{\mu\nu}$ are antisymmetric in μ and ν. This series can be written in exponential form as

$$U(\omega) = e^{\frac{i}{2}\omega_{\mu\nu}J^{\mu\nu}}.$$

Defining the operators

$$J_i := \frac{\varepsilon_{ijk}}{2} J_{jk}, \quad K_i := J^0{}_i,$$

which are, respectively, the generators of rotations and boosts along the i-axis. These operators satisfy the commutation relations

$$\begin{aligned}
\left[J_i, J_j\right] &= i\varepsilon_{ijk} J_k, \\
\left[J_i, K_j\right] &= i\varepsilon_{ijk} K_k, \\
\left[K_i, K_j\right] &= -i\varepsilon_{ijk} J_k.
\end{aligned}$$

Next, we define, in terms of the latter, the operators

$$L_i := \frac{J_i + iK_i}{2}, \qquad R_i := \frac{J_i - iK_i}{2},$$

which obey the commutation relations

$$\begin{aligned}
\left[L_i, L_j\right] &= i\varepsilon_{ijk} L_k, \\
\left[R_i, R_j\right] &= i\varepsilon_{ijk} R_k, \\
\left[L_i, R_j\right] &= 0.
\end{aligned}$$

What has been done is to split the generators of SO(3,1) into two subsets that commute with each other and which individually satisfy the commutation relations for SU(2). Now we can introduce the spinor representations.

The simplest nontrivial matrices that satisfy these commutation relations are the Pauli matrices σ_i. Now set L_i and R_i in the above equation equal to $\sigma_i/2$ and 0 respectively; then invert the relations to find J_i and K_i as

$$J_i = \frac{\sigma_i}{2}, \quad K_i = -i\frac{\sigma_i}{2}.$$

This corresponds to a left-handed spinor and, for spin ½ is designated by (0, ½). They satisfy the set of commutation relations given above for J and K. Alternatively, if we set L_i and R_i in the above equation equal to 0 and $\sigma_i/2$ respectively, and then again invert the relations to find J_i and K_i, we would obtain

$$J_i = \frac{\sigma_i}{2}, \quad K_i = i\frac{\sigma_i}{2}.$$

This corresponds to a left-handed spinor and, for spin ½ is designated by (½, 0).

If, as in the body of this essay, we consider fields and designate the right and left-handed spinor fields as Φ_R and Φ_L, and put them together into a 4-spinor, it would transform under a Lorentz transformation as

$$\Phi' = \begin{pmatrix} \Phi_R \\ \Phi_L \end{pmatrix}' = \begin{pmatrix} e^{-\frac{i}{2}(r_i + ib_i)\sigma_i} & 0 \\ 0 & e^{-\frac{i}{2}(r_i - ib_i)\sigma_i} \end{pmatrix} \begin{pmatrix} \Phi_R \\ \Phi_L \end{pmatrix},$$

where the r_i and b_i are as in the expression given above for the most general form for the $\omega^{\mu}{}_{\nu}$. This is more general than the expression given in the body of this essay because it includes rotations as well as boosts. The exponential terms in the matrix can be obtained from exponential form of $U(\omega)$ given above, by writing out explicitly $J_{\mu\nu}$ in terms of the Pauli matrices for the two choices of J_i and K_i (corresponding to (0, ½) and (½, 0), the left-handed and right handed spinor representations of the Lorentz group), and using the matrix given above for $\omega^{\mu}{}_{\nu}$ to carry out the sum indicated. Note that in obtaining the components of $J_{\mu\nu}$ involving time from the relation $K_i := J^{0}{}_{i}$, one must use $\eta_{\alpha\beta}$ with signature +2.

Appendix II: the Schwinger Term

The Schwinger term is given by

$$\mathrm{ST}(\vec{y},\vec{x}) = \left[\hat{\rho}(\vec{y}),\, \hat{\vec{J}}(\vec{x})\right]. \tag{1}$$

Taking the divergence of the Schwinger term and using the relation

$$\mathrm{i}\left[\hat{\mathrm{H}}_0,\, \hat{\rho}(\vec{x})\right] = -\nabla\cdot\, \hat{\vec{J}}(\vec{x}), \tag{2}$$

where $\hat{\mathrm{H}}_0$ is the free-field Hamiltonian when the electromagnetic 4-potential vanishes, results in

$$\nabla_{\vec{x}}\cdot \left[\hat{\rho}(\vec{y}),\, \hat{\vec{J}}(\vec{x})\right] = \left[\hat{\rho}(\vec{y}),\, \nabla\cdot\hat{\vec{J}}(\vec{x})\right] = -\mathrm{i}\left[\hat{\rho}(\vec{y}),\, \left[\hat{\mathrm{H}}_0,\, \hat{\rho}(\vec{x})\right]\right]. \tag{3}$$

Expanding the commutator on the right hand side of Eq. (3) yields the vacuum expectation value

$$i\nabla_{\vec{x}}\cdot\left\langle 0\left|\left[\hat{\rho}(\vec{y}),\hat{\vec{J}}(\vec{x})\right]\right|0\right\rangle = -\langle 0|\,\hat{\mathrm{H}}_0\,\hat{\rho}(\vec{x})\hat{\rho}(\vec{y})\,|0\rangle + \langle 0|\,\hat{\rho}(\vec{x})\hat{\mathrm{H}}_0\hat{\rho}(\vec{y})\,|0\rangle + \langle 0|\,\hat{\rho}(\vec{y})\hat{\mathrm{H}}_0\hat{\rho}(\vec{x})\,|0\rangle - \langle 0|\,\hat{\rho}(\vec{y})\hat{\rho}(\vec{x})\hat{\mathrm{H}}_0\,|0\rangle. \tag{4}$$

It is here that one makes the assumption that the vacuum is the lowest energy state. This done by writing $\hat{\mathrm{H}}_0|0\rangle = \langle 0|\hat{\mathrm{H}}_0 = 0$. As a result, Eq. (4) may be written as

$$i\nabla_{\vec{x}}\cdot\left\langle 0\left|\left[\hat{\rho}(\vec{y}),\hat{\vec{J}}(\vec{x})\right]\right|0\right\rangle = \langle 0|\,\hat{\rho}(\vec{x})\hat{\mathrm{H}}_0\hat{\rho}(\vec{y})\,|0\rangle + \langle 0|\,\hat{\rho}(\vec{y})\hat{\mathrm{H}}_0\hat{\rho}(\vec{x})\,|0\rangle. \tag{5}$$

Multiply both sides of the last equation by $f(x)f(y)$ and integrate over x and y. The right hand side of Eq. (5) becomes

$$\int d\vec{x}\,d\vec{y}\,\left\{\langle 0|\,f(\vec{x})\hat{\rho}(\vec{x})\hat{\mathrm{H}}_0 f(\vec{y})\hat{\rho}(\vec{y})\,|0\rangle + \langle 0|\,f(\vec{y})\hat{\rho}(\vec{y})\hat{\mathrm{H}}_0 f(\vec{x})\hat{\rho}(\vec{x})\,|0\rangle\right\}. \tag{6}$$

If Schwinger's "arbitrary linear functional of the charge density" is defined as

$$F = \int f(\vec{x})\hat{\rho}(\vec{x})d\vec{x} = \int f(\vec{y})\hat{\rho}(\vec{y})d\vec{y}\ , \tag{7}$$

the right hand side of Eq. (5) becomes

$$2\langle 0|\,F\hat{\mathrm{H}}_0 F\,|0\rangle = 2\sum_{m,n}\langle 0|\,F\,|m\rangle\langle m|\,\hat{\mathrm{H}}_0\,|n\rangle\langle n|\,F\,|0\rangle = 2\sum_n E_n\langle 0|\,F\,|n\rangle\langle n|\,F\,|0\rangle = 2\sum_n E_n|\langle 0|\,F\,|n\rangle|^2 > 0. \tag{8}$$

The left hand side of Eq. (8)—essentially the form used by Schwinger—is here expanded to explicitly show the non-vanishing matrix elements between the vacuum and the other states of necessarily positive energy. This shows that if the vacuum is assumed to be the lowest energy state, the Schwinger term cannot vanish, and the theory is not gauge invariant.

For the sake of completeness, it is readily shown that the left side of Eq. (5) becomes

$$i\int \nabla_{\vec{x}}\cdot\left\langle 0\left|\left[\hat{\rho}(\vec{y}),\hat{\vec{J}}(\vec{x})\right]\right|0\right\rangle f(\vec{x})f(\vec{y})\mathrm{d}\vec{x}\mathrm{d}\vec{y} = i\,\langle 0\,|\,[\partial_t F, F]\,|0\rangle, \tag{9}$$

so that combining Eqs. (8) and (9) yields a somewhat more explicit form of the result given by Schwinger,

$$i\,\langle 0\,|\,[\partial_t F, F]\,|0\rangle = 2\sum_n E_n|\langle 0|\,F\,|n\rangle|^2 > 0. \tag{10}$$

A Few Reference Books

The following list of references is by no means representative of the vast array of books and articles published on the subject matter discussed above. Of the many fine books that I consulted, these simply represent a few that I found particularly well written. The scientific literature is far too vast to be referred to here and key references may be found in the books below.

Aitchison, I. J. R., and Hey, A. J. G., *Gauge Theories in Particle Physics* (Institute of Physics Publishing 1989).

Aitchison, I. J. R., *An Informal Introduction to Gauge Field Theories* (Cambridge University Press 1984).

Halzen, F. and Martin, A. D. Quarks and Leptons: *An Introductory Course in Modern Particle Physics* (John Wiley and Sons 1984).

Moriyasu, K., *An Elementary Primer for Gauge Theory* (World Scientific 1983)

Ryder, L. H., *Quantum Field Theory* (Cambridge University Press 1996).

Sternberg, S., *Group Theory and Physics* (Cambridge University Press 1994).

Weyl, H., *Symmetry* (Princeton University Press 1952).