## 21.3 The Electroweak Theory

The most important application of spontaneously broken gauge theories has been to the theory of weak and electromagnetic interactions.[3] Weak interactions at low energies are well described by an effective Lagrangian given by a sum of products of vector (including axial-vector) currents, as in Eq. (19.4.22). This suggests that these interactions may like electromagnetism be described by some sort of gauge theory. In order to insure the separate conservation of electronic-type and muonic-type leptons and baryons (or quarks) we may guess that the known electronic-type and muonic-type leptons and the quarks all form separate representations of the gauge group. With this assumption, there are only a few possibilities for the structure of the gauge group.

Let's first consider the electronic-type lepton fields. As far as we know, these consist only of the left- and right-handed parts of the electron field $e$:

$$e_L = \tfrac{1}{2}(1+\gamma_5)e\,, \qquad e_R = \tfrac{1}{2}(1-\gamma_5)e\,, \tag{21.3.1}$$

and a purely left-handed electron-neutrino field $\nu_{eL}$:

$$\gamma_5 \nu_{eL} = \nu_{eL}\,. \tag{21.3.2}$$

The fields in any representation of the gauge group must all have the same Lorentz-transformation properties, so the representations of the gauge group here divide* into a left-handed doublet ($\nu_{eL}$, $e_L$) and a right-handed singlet $e_R$. The largest possible gauge group is then

$$SU(2)_L \times U(1)_L \times U(1)_R\,,$$

under which the fields transform as

$$\delta\begin{pmatrix}\nu_e\\ e\end{pmatrix} = i\left[\vec{\epsilon}\cdot\vec{t} + \epsilon_L t_L + \epsilon_R t_R\right]\begin{pmatrix}\nu_e\\ e\end{pmatrix}, \tag{21.3.3}$$

where the generators are

$$\vec{t} = \frac{g}{4}(1+\gamma_5)\left\{\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right\}, \tag{21.3.4}$$

$$t_L \propto (1+\gamma_5)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{21.3.5}$$

$$t_R \propto (1-\gamma_5)\,, \tag{21.3.6}$$

---

* If we allow gauge couplings that change electron-type lepton number, then it is possible to include the left-handed field $\overline{e_R}$ along with $\nu_{eL}$ and $e_L$ in a representation of the gauge group. This was the basis for an early $SO(3)$ variant[6] of the electroweak theory, which has since been ruled out by experiment.

with $g$ a constant to be chosen later. It will be convenient instead of $t_L$ and $t_R$ to consider the generators

$$y \equiv g' \left[ \left( \frac{1+\gamma_5}{4} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \left( \frac{1-\gamma_5}{2} \right) \right] \tag{21.3.7}$$

and

$$n_e \equiv g'' \left[ \left( \frac{1+\gamma_5}{2} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \left( \frac{1-\gamma_5}{2} \right) \right] , \tag{21.3.8}$$

where $g'$ and $g''$ are constants like $g$ to be chosen later. The generator $y$ appears along with $t_3$ in a linear combination that plays a special role in physics; it is the electric charge

$$q = e \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} = \frac{e}{g} t_3 - \frac{e}{g'} y . \tag{21.3.9}$$

Also, $n_e$ is the electron-type lepton number. We want to include both charge-changing weak interactions (like beta decay) and electromagnetism in our theory, so we will assume that there are gauge fields $\vec{A}^\mu$ and $B^\mu$ coupled to $\vec{t}$ and $y$. In addition, we may or may not want to include a gauge field coupled to the one remaining independent linear combination of $t_L$ and $t_R$, which can be taken as the electron-type lepton number (21.3.8). There are very stringent limits[7] on the long-range forces that would be produced by a massless gauge field coupled to $n_e$, so in order to include in our theory a gauge field coupled to $n_e$ with a strength $g''$ comparable to the weak and electromagnetic interactions, we would have to assume that this gauge symmetry is spontaneously broken.** However, there is no experimental evidence for the weak interaction that would be produced by such a gauge coupling (and plenty of evidence by now against it) so we shall simply exclude $n_e$ from the generators of the gauge group. The gauge group is then[8]

$$G = SU(2)_L \times U(1) \tag{21.3.10}$$

with generators $\vec{t}, y$ given by Eqs. (21.3.4) and (21.3.7) respectively. The coupling constants $g$ and $g'$ are to be adjusted so that the gauge fields $\vec{A}^\mu$ and $B^\mu$ coupled to these generators are canonically normalized. The most general gauge-invariant and renormalizable Lagrangian that involves just

** Note that this is possible without violating the global conservation law of electronic lepton conservation. We would have to assume that the Lagrangian is invariant under both a global phase transformation acting only on electron-type lepton fields, and also a local phase transformation that acts on electron-type lepton fields as well as on some scalar field that does not interact with leptons. The vacuum expectation value of this scalar would break the local symmetry, giving the gauge boson coupled to $n_e$ a mass, without breaking the global symmetry.

these gauge fields and electronic leptons is then

$$\mathscr{L}_{YM} + \mathscr{L}_e = -\tfrac{1}{4}\Big(\partial_\mu \vec{A}_\nu - \partial_\nu \vec{A}_\mu + g\vec{A}_\mu \times \vec{A}_\nu\Big)^2 - \tfrac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2$$
$$-\bar{\ell}(\not\partial - i\not\vec{A}\cdot\vec{t}_L - i\not B\,y)\ell\,. \qquad (21.3.11)$$

(We here use the fact that the structure constants of $SU(2)_L$ and $U(1)$ are $C_{ijk} = -i\,g\,\epsilon_{ijk}$ and zero, respectively.)

Of course, of the four gauge fields coupled to $\vec{t}$ and $y$, only one linear combination, the electromagnetic field $A^\mu$, is actually massless. We therefore must assume that $SU(2)_L \times U(1)$ is spontaneously broken to a subgroup $U(1)_{em}$, with generator given by the charge (21.3.7). The details of the symmetry-breaking mechanism will be considered a little later. However, whatever this mechanism may be, we know that the canonically normalized vector fields corresponding to particles of spin one and definite mass consist of one field of charge $+e$ with mass $m_W$:

$$W^\mu = \frac{1}{\sqrt{2}}(A_1^\mu + i\,A_2^\mu)\,, \qquad (21.3.12)$$

another of charge $-e$ and the same mass:

$$W^{\mu *} = \frac{1}{\sqrt{2}}(A_1 - i\,A_2^\mu)\,, \qquad (21.3.13)$$

and two electrically neutral fields of mass $m_Z$ and zero respectively, given by orthonormal linear combinations of $A_3^\mu$ and $B^\mu$:

$$Z^\mu = \cos\theta\, A_3^\mu + \sin\theta\, B^\mu\,, \qquad (21.3.14)$$

$$A^\mu = -\sin\theta\, A_3^\mu + \cos\theta\, B^\mu\,, \qquad (21.3.15)$$

or equivalently

$$A_3^\mu = \cos\theta\, Z^\mu - \sin\theta\, A^\mu\,, \qquad (21.3.16)$$

$$B^\mu = \sin\theta\, Z^\mu + \cos\theta\, A^\mu\,. \qquad (21.3.17)$$

According to the general result (21.1.11)–(21.1.12), the generator of the unbroken symmetry, which is here electromagnetic gauge invariance, is given by a linear combination of generators in which the coefficients are the same as the coefficients of the corresponding massless field in the expansion of the canonically normalized gauge fields coupled to these generators. Inspecting Eqs. (21.3.16) and (21.3.17) shows that

$$q = -\sin\theta\, t_3 + \cos\theta\, y\,. \qquad (21.3.18)$$

Comparing this with Eq. (21.3.9) gives then

$$g = -e/\sin\theta\,, \qquad g' = -e/\cos\theta\,. \qquad (21.3.19)$$

The complete lepton–gauge boson coupling can be expressed in terms of the couplings g and g′:

$$
\begin{aligned}
i\mathscr{L}'_e &= -\overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}}\left[\sum_\alpha \not{A}_\alpha t_\alpha\right]\begin{pmatrix} \nu_e \\ e \end{pmatrix} \\
&= -\overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}}\left[\frac{1}{\sqrt{2}}\ \not{W}(t_{1L} - it_{2L}) + \frac{1}{\sqrt{2}}\ \not{W}^*(t_{1L} + it_{2L})\right. \\
&\qquad \left. + \not{Z}\,(t_{3L}\cos\theta + y\sin\theta) + \not{A}(-t_{3L}\sin\theta + y\cos\theta)\right]\begin{pmatrix} \nu_e \\ e \end{pmatrix} \\
&= \frac{g}{\sqrt{2}}\left(\bar{e}\ \not{W}\left(\frac{1+\gamma_5}{2}\right)\nu_e\right) + \frac{g}{\sqrt{2}}\left(\bar{\nu}_e\ \not{W}^*\left(\frac{1+\gamma_5}{2}\right)e\right) \\
&\quad -\frac{1}{2}\sqrt{g^2+g'^2}\,\bar{\nu}_e\ \not{Z}\left(\frac{1+\gamma_5}{2}\right)\nu_e + \frac{(g^2-g'^2)}{2\sqrt{g^2+g'^2)}}\ \bar{e}\ \not{Z}\left(\frac{1+\gamma_5}{2}\right)e \\
&\quad +g'\bar{e}\ \not{Z}\left(\frac{1-\gamma_5}{2}\right)e - e\,(\bar{e}\not{A}\,e)\,.
\end{aligned} \tag{21.3.20}
$$

To complete the theory, we must now make some assumption about the mechanism of symmetry breaking. We want this mechanism to give masses not only to the $W^\pm$ and $Z^0$, but to the electron as well. Now, the only way that this is possible in a renormalizable weakly-coupled theory is to have a scalar field coupled without derivatives to $\bar{\ell}_R$ and $\ell_L$ (and also $\bar{\ell}_L$ and $\ell_R$ ). Then $SU(2)_L \times U(1)$ invariance requires that the scalar be an $SU(2)_L$ doublet like $\ell_L$, but with a shifted value of $y$ and hence of $q$. We thus assume a 'Yukawa' coupling

$$
\mathscr{L}_{\phi e} = -G_e\overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R \ + \text{H.c.}\,, \tag{21.3.21}
$$

where $(\phi^+, \phi^0)$ is a doublet, on which the $SU(2) \times U(1)$ generators are represented by the matrices:

$$
\vec{t}^{(\phi)} = \frac{g}{2}\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right\}, \tag{21.3.22}
$$

$$
y^{(\phi)} = -\,g'/2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{21.3.23}
$$

so that the charge matrix is

$$
q^{(\phi)} = \frac{e}{g}\,t_3^{(\phi)} - \frac{e}{g'}\,y^{(\phi)} = e\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{21.3.24}
$$

It is possible that there are other scalar multiplets in the theory, but for the moment let's suppose that this is the only one.

We must add a gauge-invariant term involving scalar and gauge fields to the Lagrangian. The most general form consistent with $SU(2) \times U(1)$

gauge invariance, Lorentz invariance, and renormalizability is:

$$\mathscr{L}_\phi = -\frac{1}{2}\left|(\partial_\mu - i\vec{A}_\mu\cdot\vec{t}^{(\phi)} - i\,B_\mu\, y^{(\phi)})\phi\right|^2 - \frac{\mu^2}{2}\,\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2\,, \quad (21.3.25)$$

where $\lambda > 0$, and

$$\phi \equiv \begin{pmatrix}\phi^+ \\ \phi^0\end{pmatrix}. \quad (21.3.26)$$

For $\mu^2 < 0$, there is a tree-approximation vacuum expectation value at the stationary point of the Lagrangian

$$\langle\phi\rangle^\dagger\langle\phi\rangle \equiv v^2 = |\mu^2|/\lambda\,. \quad (21.3.27)$$

We can always perform an $SU(2)\times U(1)$ gauge transformation to a unitarity gauge, in which $\phi^+ = 0$ and $\phi^0$ is Hermitian, with a positive vacuum expectation value. (This is why we normalized the complex doublet $\phi$ so that an unconventional factor $\frac{1}{2}$ appears in the kinetic term in Eq. (21.3.25); $\text{Re}\,\phi^0$ is the only physical scalar field, and Eq. (21.3.25) makes this a canonically normalized field.) In unitarity gauge the vacuum expectation values of the components of $\phi$ are

$$\langle\phi^+\rangle = 0\,, \qquad \langle\phi^0\rangle = v > 0\,. \quad (21.3.28)$$

The scalar Lagrangian (21.3.25) then yields a vector meson mass term

$$-\frac{1}{2}\left|\left(\vec{A}_\mu\cdot\vec{t}^{(\phi)} + B_\mu y^{(\phi)}\right)\langle\phi\rangle\right|^2 = -\frac{1}{2}\left|\left(\frac{g}{2}\vec{A}_\mu\cdot\vec{\tau} - \frac{g'}{2}B_\mu\right)\begin{pmatrix}0\\ v\end{pmatrix}\right|^2$$
$$= -\frac{v^2 g^2}{4}\,W_\mu^\dagger W^\mu - \frac{v^2}{8}(g^2 + g'^2)\,Z_\mu Z^\mu\,. \quad (21.3.29)$$

We see that as expected, the photon mass is zero, while the $W^\pm$ and $Z^0$ have the masses

$$m_W = \frac{v|g|}{2}\,, \qquad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2}\,. \quad (21.3.30)$$

Also, from Eqs. (21.3.21) and (21.3.28) we see that the electron is given a lowest-order mass

$$m_e = G_e v\,. \quad (21.3.31)$$

It is difficult to study reactions among electron-type leptons alone, though by now there are data on scattering processes like $\bar{\nu}_e + e^- \to \bar{\nu}_e + e^-$. For high precision data we have to consider reactions that also involve at least involve muonic-type leptons, such as the well-studied process of muon decay, $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$. It is trivial to extend the above model to include muon-type leptons — just add to the Lagrangian terms $\mathscr{L}_\mu$

and $\mathscr{L}_{\phi\mu}$, like the last terms in Eqs. (21.3.11) and (21.3.21), with the fields $e$ and $\nu_e$ replaced with the muon and muon-neutrino fields $\mu^-$ and $\nu_\mu$, and with $G_e$ replaced with $G_\mu = G_e(m_\mu/m_e)$. Inspection of (21.3.20) and the corresponding term with $e$ and $\nu_e$ replaced with $\mu$ and $\nu_\mu$ shows that $W$ exchange between low energy, $e$-type and $\mu$-type leptons produces the effective interaction

$$\left(\frac{g}{\sqrt{2}}\right)^2 \frac{1}{m_W^2} \left(\bar{e}\gamma^\lambda \left(\frac{1+\gamma_5}{2}\right) \nu_e\right) \left(\bar{\nu}_\mu \gamma_\lambda \left(\frac{1+\gamma_5}{2}\right) \mu\right) + \text{H.c.} \tag{21.3.32}$$

This may be compared with the interaction of the effective '$V-A$' theory which is known to give a good description of muon decay

$$\frac{G_F}{\sqrt{2}} \left(\bar{e}\gamma^\lambda (1+\gamma_5) \nu_e\right) \left(\bar{\nu}_\mu \gamma_\lambda (1+\gamma_5) \mu\right) + \text{H.c.} \tag{21.3.33}$$

Here $G_F$ is the conventional Fermi coupling constant, known from the muon decay rate to have the value $G_F = 1.16639(2) \times 10^{-5}\,\text{GeV}^{-2}$. Comparing these two expressions, we find

$$g^2/m_W^2 = 4\sqrt{2}\, G_F\,. \tag{21.3.34}$$

This allows an immediate determination of the vacuum expectation value $v$, given by Eq. (21.3.30) as

$$v = \frac{2m_W}{g} = \frac{1}{2^{1/4} G_F^{1/2}} = 247\,\text{GeV}\,. \tag{21.3.35}$$

Also, Eq. (21.3.31) shows that $G_e$ has the very small value

$$G_e = \frac{0.511\,\text{MeV}}{247\,\text{GeV}} = 2.07 \times 10^{-6}\,. \tag{21.3.36}$$

From Eq. (21.3.30) we see that $m_Z > m_W$. We cannot use Eq. (21.3.30) to determine the actual values of $m_Z$ and $m_W$ without knowing something about $g$ and $g'$. Using Eqs. (21.3.30) and (21.3.19), we can express $m_Z$ and $m_W$ in terms of the electroweak mixing angle $\theta$:

$$m_W = \frac{ev}{2|\sin\theta|} = \frac{37.3\,\text{GeV}}{|\sin\theta|}\,,$$

$$m_Z = \frac{ev}{2|\sin\theta||\cos\theta|} = \frac{74.6\,\text{GeV}}{|\sin 2\theta|}\,.$$

These are the original results obtained in Ref. 3. Of course, there are radiative corrections of all sorts, most of which depend on details of the theory that have not yet been specified in this section. But there is one particularly large radiative correction that can be readily calculated without further information. The above values for $m_W$ and $m_Z$ were calculated using the conventionally defined electronic charge for $e$. However, as explained in Section 18.2, this is not precisely the appropriate value to use in calculations of processes at energies $E \gg m_e$; we should instead use the electric charge $e_\mu$ defined at a sliding scale $\mu$ comparable to the energies of interest. For $\mu$ of the order of 90 GeV the effective fine structure constant $e_\mu^2/4\pi$ is about 1/129 (and quite insensitive to the precise value of $\mu$), so the above values for $m_W$ and $m_Z$ should be multiplied with $\sqrt{137/129}$, giving

$$m_W = \frac{38.4 \text{ GeV}}{|\sin\theta|}\,, \tag{21.3.37}$$

$$m_Z = \frac{76.9 \text{ GeV}}{|\sin 2\theta|}\,. \tag{21.3.38}$$

Whatever the value of $\theta$, these masses are too large for there to have been any hope of detecting the $W$ or $Z$ in the 1960s or early 1970s. Experimental evidence for the electroweak theory had to come instead from the discovery of the new class of weak interactions predicted by the theory, the neutral current processes produced by $Z^0$ exchange.[9] The first observation of a neutral current process was the 1973 bubble chamber detection of the purely leptonic process of $\nu_\mu$–$e^-$ elastic scattering.[10] Although these processes are easy to deal with theoretically, the frequency of events is relatively low, because the cross section is proportional to the square of the center-of-mass energy.[†] It was years before the purely leptonic neutral current reactions could be used to give a reasonably precise value for the parameter $\sin^2\theta$. By 1994, the study of purely leptonic neutral current processes like $\nu_\mu + e^- \to \nu_\mu + e^-$ and $\bar{\nu}_\mu + e^- \to \bar{\nu}_\mu + e^-$ had yielded the value $0.222 \pm 0.011$, which would give $m_W = 81.5$ GeV and $m_Z = 92.5$ GeV.

Even before the discovery of neutral currents, the electroweak theory had been extended to the weak and electromagnetic interactions of hadrons with each other and with leptons. By the mid-1960s, it had become understood that weak interaction processes in which charge is exchanged between leptons and hadrons are well described at low energy by the effective Lagrangian

$$\frac{G_F}{\sqrt{2}}\left[\bar{e}\gamma_\lambda(1+\gamma_5)\nu_e + \bar{\mu}\gamma_\lambda(1+\gamma_5)\nu_\mu\right] J^\lambda \;+\; \text{H.c.}\,, \tag{21.3.39}$$

where $J^\lambda$ is an hadronic current. Within the quark model, the commutation and conservation properties of $J^\lambda$ allowed it to be identified with the quark

[†] The cross section is proportional to $G_F^2$, so in order to have the dimensions of energy$^{-2}$, it must also be proportional to some energy squared. Where the center-of-mass energy is much larger than the electron mass, it is the only energy that can appear in this formula.

current

$$J^\lambda = \bar{u}\gamma^\lambda(1+\gamma_5)d\cos\theta_c + \bar{u}\gamma^\lambda(1+\gamma_5)s\,\sin\theta_c\,. \tag{21.3.40}$$

Here $u, d, s$ are the fields of the up, down, and strange quarks, and $\theta_c$ is another angle, known as the Cabibbo angle.[11] Experiments on processes like $O^{14} \to N^{14*} + e^+ + \nu_e$ and $K^+ \to \pi^0 + e^+ + \nu_e$ confirm that $G_F$ has very nearly the same value as that measured in the purely leptonic process $\mu^+ \to \bar{\nu} + e^+ + \nu_e$, and give for $\theta_c$ the value[12] $\sin\theta_c = 0.220 \pm 0.003$. We naturally conclude that the quarks provide another $SU(2)\times U(1)$ doublet

$$\mathcal{Q} = \left(\frac{1+\gamma_5}{2}\right)\begin{bmatrix} u \\ d\cos\theta_c + s\,\sin\theta_c \end{bmatrix}, \tag{21.3.41}$$

as well as right-handed singlets, with $y$ values adjusted to give the quark charges $2e/3$ and $-e/3$. By itself this would lead to a serious difficulty. The $Z^0$ boson interacts with the quark neutral current

$$\sum_{\mathcal{Q}} \bar{\mathcal{Q}}\gamma^\mu(t_{3L}\cos\theta + y\sin\theta)\mathcal{Q} = \sum_{\mathcal{Q}} \bar{\mathcal{Q}}\gamma^\mu(t_{3L}\sec\theta + q\tan\theta)\mathcal{Q}\,, \tag{21.3.42}$$

with the sum running over all quark doublets $\mathcal{Q}$ like (21.3.41). The charge matrix $q$ is diagonal in quark flavors, but if (21.3.41) were the only quark doublet then the term involving the matrix $t_{3L}$ would contain cross terms proportional to $\bar{s}\gamma^\mu(1+\gamma_5)d$ and $\bar{d}\gamma^\mu(1+\gamma_5)s$, leading to effective $Z$ exchange interactions like $s + \bar{d} \leftrightarrow d + \bar{s}$ and $s + \bar{d} \leftrightarrow \mu^+ + \mu^-$ with the strength of ordinary first-order weak interactions. Such effects would lead to rates for processes like $K^0$–$\overline{K^0}$ oscillations and $K^0 \to \mu^+ + \mu^-$ many orders of magnitude greater than observed. Also, even without neutral current terms in the Lagrangian, the one-loop diagrams involving the interaction (21.3.39) with the charged current (21.3.40) would lead to an effective interaction $s + \bar{d} \to d + \bar{s}$ which is smaller than an ordinary first-order weak interaction only by a factor of order $\alpha/2\pi$, leading to a rate for $K^0$–$\overline{K^0}$ oscillations that is still much too large. In order to avoid this last difficulty, it was proposed[13] that there is another term in $J^\lambda$; in modern notation,

$$\bar{c}\gamma^\lambda(1+\gamma_5)\,[-d\,\sin\theta_c + s\cos\theta_c]\,, \tag{21.3.43}$$

where $c$ is a fourth quark, like $u$ with charge $2e/3$. Adding (21.3.43) to (21.3.40), the charged current may be written

$$J^\lambda = (\bar{u}\cos\theta_c - \bar{c}\sin\theta)\gamma^\lambda(1+\gamma_5)d + (\bar{u}\sin\theta_c + \bar{c}\cos\theta_c)\gamma^\lambda(1+\gamma_5)s\,.$$

The only reason that the interactions of the $W$ with this current do not conserve strangeness is that the $c$ and $u$ have different masses, leading to transitions between $u\cos\theta_c - c\sin\theta$ and $u\sin\theta_c + c\cos\theta_c$. But this means that the loop diagrams for the effective interaction $s + \bar{d} \to d + \bar{s}$ are suppressed by additional factors (since $m_u \ll m_c$) of $m_c^2/m_W^2$, bringing the rate for $K^0$–$\overline{K^0}$ oscillations into agreement with experiment.

It was subsequently noted[14] that this also solves the problem of the strangeness changing $Z^0$ interactions. In the context of the $SU(2) \times U(1)$ gauge theory the combination $-d_L \sin\theta_c + s_L \cos\theta_c$ cannot be a singlet, but must be part of another doublet

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} c \\ -d\,\sin\theta_c + s\,\cos\theta_c \end{bmatrix} . \tag{21.3.44}$$

Including this doublet in the weak neutral current (21.3.42), the strangeness non-conserving terms proportional to $\bar{s}\gamma^\mu(1+\gamma_5)d$ and $\bar{d}\gamma^\mu(1+\gamma_5)s$ cancel, removing the problem of excessive $Z$ exchange contributions to processes like $K^0$–$\overline{K^0}$ oscillations and $K^0 \to \mu^+ + \mu^-$. Particles containing the $c$ quark in a $c$–$\bar{c}$ bound state were discovered[15] in 1974, and indicated a mass $m_c \approx 1.5$ GeV.[††] This completed two generations of quarks and leptons: a $(u, d)$ quark doublet mixed with a $(c, s)$ quark doublet, together with two lepton doublets $(\nu_e, e)$ and $(\nu_\mu, \mu)$.

The first sign of a third generation was the discovery of a third charged lepton,[16] the $\tau$. Later a fifth quark type, the $b$, was discovered,[17] with charge $-e/3$ and a mass of about 4.5 GeV. A sixth, the $t$ with charge $2e/3$, then became theoretically necessary, and after a long interval it too was discovered,[18] with a mass quoted in 1995 as $181 \pm 12$ GeV.[19] Today the hadronic current $J^\lambda$ in (21.3.39) is expressed as

$$J^\lambda = \overline{\begin{bmatrix} u \\ c \\ t \end{bmatrix}} \gamma^\lambda(1+\gamma_5)\,V \begin{bmatrix} d \\ s \\ b \end{bmatrix} , \tag{21.3.45}$$

where $V$ is an incompletely known $3 \times 3$ unitary matrix, known as the Kobayashi–Maskawa matrix.[20] In the $SU(2) \times U(1)$ gauge theory, this means that there are three quark doublets:

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} u \\ V_{ud}d + V_{us}s + V_{ub}b \end{bmatrix} , \tag{21.3.46}$$

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} c \\ V_{cd}d + V_{cs}s + V_{cb}b \end{bmatrix} , \tag{21.3.47}$$

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} t \\ V_{td}d + V_{ts}s + V_{tb}b \end{bmatrix} . \tag{21.3.48}$$

[††] We do not observe quarks in isolation, so their masses are not precisely defined. The mass of the $c$ quark quoted here is roughly half the mass of the $J$–$\psi$ particle, interpreted as a $c$ $\bar{c}$ bound state. The $b$ and $t$ quarks are so heavy that their masses can be taken from the masses of the hadrons containing them with little ambiguity.

It is important to recognize that this is just what we should naturally expect on general grounds for three quark doublets. The most general renormalizable $(SU(3)\times SU(2)\times U(1))$-invariant interactions of the scalar doublets $\phi_n$ with the quarks must in general take the form

$$\mathscr{L}_\phi = -\sum_{ijn} G^n_{ij}\overline{\begin{pmatrix} U_{iL} \\ D_{iL}\end{pmatrix}}\cdot\begin{pmatrix}\phi^0_n \\ \phi^-_n\end{pmatrix} U_{jR}$$

$$-\sum_{ijn} H^n_{ij}\overline{\begin{pmatrix} U_{iL} \\ D_{iL}\end{pmatrix}}\cdot\begin{pmatrix}-\phi^{-\dagger}_n \\ \phi^{0\dagger}_n\end{pmatrix} D_{jR} + \text{H.c.}\,, \qquad (21.3.49)$$

where $U_i$ and $D_i$ with $i = 1, 2, 3$ are three independent quark fields of charge $2e/3$ and $-e/3$, respectively, $L$ and $R$ denote the left- and right-handed parts of the quark fields, and $G^n_{ij}$ and $H^n_{ij}$ are unknown constants. The vacuum expectation values of the neutral scalars then produce a quark mass term

$$\mathscr{L}_m = -\sum_{ij}\overline{U_{iL}}\, m^U_{ij}\, U_{jR} - \sum_{ij}\overline{D_{iL}}\, m^D_{ij}\, D_{jR} + \text{H.c.}\,, \qquad (21.3.50)$$

where

$$m^U_{ij} = \sum_n G^n_{ij}\langle\phi^0_n\rangle_{\text{VAC}}\,, \qquad m^D_{ij} = \sum_n H^n_{ij}\langle\phi^0_n\rangle^*_{\text{VAC}}\,. \qquad (21.3.51)$$

The matrices $m^U_{ij}$ and $m^D_{ij}$ are not constrained in any way, and in particular may be complex and non-diagonal, in which case parity- and flavor-non-conserving terms appear in $\mathscr{L}_m$. But we can introduce new quark fields $U'_R = A^U_R U_R$, $U'_L = A^U_L U_L$, $D'_R = A^D_R D_R$, $D'_L = A^D_L D_L$, where the $A$s are $3\times 3$ matrices constrained only by the condition that they must be unitary in order to preserve the form of the kinematic term (19.4.1). Then the mass term (21.3.50) takes the same form when rewritten in terms of the primed quark fields, but with the matrices $m^U$ and $m^D$ replaced with

$$m^{U\prime} = A^U_L m^U A^{U\dagger}_R\,, \qquad m^{D\prime} = A^D_L m^D A^{D\dagger}_R\,. \qquad (21.3.52)$$

Now it is a general theorem that for any matrix $m$, it is always possible to choose unitary matrices $A$ and $B$ such that $AmB$ is real and diagonal. (Use the polar decomposition theorem to write $m = HU$, where $H$ is Hermitian and $U$ is unitary, and choose $A = S^\dagger$ and $B = U^\dagger S$, where $S$ is the unitary matrix that diagonalizes $H$.) We can therefore choose the $A$s so that $m^{U\prime}$ and $m^{D\prime}$ are real and diagonal, in which case the quark fields $u$, $c$, $t$, $d$, $s$, and $b$ are to be identified with the components of $U'_L + U'_R$ and $D'_L + D'_R$. The weak doublets are now written as

$$Q_{iL} = \begin{pmatrix} (A^{U-1}_L U'_L)_i \\ (A^{D-1}_L D'_L)_i \end{pmatrix}\,,$$

but we can just as well take the doublets as linear combinations $A_L^U Q_L$ that have charge $2e/3$ quarks of definite mass $u$, $c$, $t$ as their top component, in which case these doublets take the form (21.3.46)–(21.3.48), with

$$V = A_L^U A_L^{D-1} \,. \tag{21.3.53}$$

Within 90% confidence limits, the best present (1995) values for the absolute values of the elements of the Kobayashi–Maskawa matrix are[20a]

$$\begin{pmatrix} 0.9745 \text{ to } 0.9757 & 0.219 \text{ to } 0.224 & 0.002 \text{ to } 0.005 \\ 0.218 \text{ to } 0.224 & 0.9736 \text{ to } 0.9750 & 0.036 \text{ to } 0.046 \\ 0.004 \text{ to } 0.014 & 0.034 \text{ to } 0.046 & 0.9989 \text{ to } 0.9993 \end{pmatrix} ,$$

with rows labelled $u$, $c$, and $t$, and columns labelled $d$, $s$, and $b$.

If there were only two quark doublets formed from $u$, $d$, $c$, and $s$ quarks, it would be possible to choose the phases of the quark fields so that all $V_{ij}$ are real,‡ so that the $V$ matrix is orthogonal, and the doublets (21.3.46) and (21.3.47) (with $b$ omitted) take the form (21.3.41) and (21.3.44), respectively. In this case the gauge interactions would automatically conserve T and CP. The great importance of the third generation is that it is no longer always possible to choose quark phases so that the $V$ matrix is real, and therefore the gauge interactions can violate T and CP conservation. But for unknown reasons the elements $V_{ub}$, $V_{cb}$, $V_{td}$, and $V_{ts}$ that connect the third generation with the first two are all quite small, so the physics of the first two generations is hardly affected by the presence of the third, which explains in a more-or-less natural way why the Cabibbo assumption (21.3.40) works so well and why the violation of T and CP conservation is so weak. T and CP conservation can also be violated by scalar boson interactions if there are two or more scalar doublets;[20b] here the violation of T and CP conservation is expected to be weak because the scalar doublets couple weakly to light quarks. It is still unknown which of these mechanisms is responsible for the observed violation of T and CP conservation in $K_2^0$ decay, discussed in Section 3.3.

Neutral current processes involving hadrons, such as neutrino–nucleon deep-inelastic scattering, were discovered in 1973,[21] shortly after the detection of the purely leptonic process $\nu_\mu + e \to \nu_\mu + e$. Because of the much greater mass of the target particle here, it became possible before long to observe large numbers of events, and use them to confirm the electroweak theory and measure its parameters. Additional information on lepton–hadron neutral current interactions came from the observation of parity violation in atomic physics. By 1983 all direct measurements of

‡ Adjust the phases of $d$ and $s$ so that $V_{ud}$ and $V_{us}$ are real. Unitarity then requires that $V_{cd}$ and $V_{cs}$ have the same phase, which can be eliminating by adjusting the phase of $c$.

$\sin^2\theta$ had become consistent, and gave a combined value $\sin^2\theta = 0.23$, yielding the predictions $m_W = 80.1\,\text{GeV}$ and $m_Z = 91.4\,\text{GeV}$. Then in 1983 the $W$ was discovered, with the $Z$ following soon after.[22] Their measured masses are now (in 1995)

$$m_W = 80.410 \pm 0.180\ \text{GeV}^{23}\,, \qquad m_Z = 91.1887 \pm 0.0022\ \text{GeV}^{24}\,,$$

in satisfactory agreement with the predictions of the electroweak theory.

The very great accuracy of the measurement of the $Z$ mass, which has been achieved by tuning the energy of $e^+$–$e^-$ collisions to the $Z$ resonance at LEP (CERN's Large Electron Positron collider) and the SLC (Stanford Linear Collider), has changed the way that electroweak data is analyzed. Instead of comparing predictions of $W$ and $Z$ masses with observed values, the $Z$ mass is taken as an experimental input, along with the Fermi coupling constant $G_F = 1.16639(2) \times 10^{-5}\ \text{GeV}^{-2}$ taken from the rate of muon decay (including radiative corrections to order $\alpha$), and the fine structure constant $\alpha(m_Z) = (128.87 \pm 0.12)^{-1}$, extrapolated from low energy measurements as described in Section 18.2. In this way $\sin^2\theta$ becomes a derived quantity; if defined by Eq. (21.3.38), it takes the value $\sin^2\theta = 0.2312 \pm 0.003$. With these inputs, the electroweak theory can be used to make predictions of other quantities like $m_W$ with sufficient precision that it becomes necessary to take electroweak radiative corrections into account.[25] In one-loop order these radiative corrections involve the masses of the $t$ quark and scalar ('Higgs') boson, and thus can be used to estimate these masses. For instance, before the top quark was discovered the agreement between theory and experiment set bounds on these radiative corrections which implied a top quark mass in the range 130–200 GeV,[26] in agreement with the value subsequently found experimentally. The $W$ mass is predicted (in 1994) to be 80.29 GeV, with an uncertainty of $\pm 0.02$ GeV from uncertainties in the inputs $m_Z$, $G_F$, and $\alpha(m_Z)$, and an uncertainty of $\pm 0.11$ GeV from the range of possible values of $m_t$ and $m_{\text{Higgs}}$. One 1995 study[27] concludes that $m_{\text{Higgs}} < 225$ GeV. The precise measurement of $m_W$ expected at the LEP 2 electron–positron collider at CERN will allow a useful estimate of $m_{\text{Higgs}}$.

* * *

The most general renormalizable Lagrangian with the field content and $SU(3) \times SU(2) \times U(1)$ gauge symmetries of the electroweak theory automatically respects baryon and lepton conservation. This is obviously true for the gauge interactions and bare mass terms, because the quarks, antiquarks, leptons, and antileptons all belong to distinct representations of $SU(3) \times SU(2) \times U(1)$. With scalars all belonging to $SU(3)$ neutral $SU(2)$ doublets with $U(1)$ quantum number $\pm 1/2$, the only renormalizable interactions of scalars with fermions and/or antifermions are with quark–

antiquark and lepton–antilepton pairs, which of course conserve baryon and lepton number. (In much the same way, one can see that the charged hadronic currents with which leptons interact are necessarily linear combinations of the currents associated with the spontaneously broken $SU(3) \times SU(3)$ symmetry described in Section 19.7, as assumed without knowing the explanation in the original work on this broken symmetry.)

These results depend critically on the assumption that the standard model is renormalizable. But as we have repeatedly emphasized, the renormalizable Lagrangian of the standard model is expected to be accompanied with non-renormalizable terms of dimensionality $d > 4$, suppressed by $4 - d$ powers of some very large mass $M$. The leading corrections to the predictions of the renormalizable standard model come from terms with the smallest possible dimensionality greater than four.

The only Lorentz-invariant terms of dimensionality five that can be constructed out of the fermion and other fields of the standard model are at most bilinear in fermion fields and also contain either two scalars, or one scalar and one gauge-invariant derivative, or no scalars and two gauge-invariant derivatives (including their commutator, a field strength tensor). Color $SU(3)$ invariance requires that the fermion fields in such an interaction appear in either a quark–antiquark bilinear or a pair of lepton and/or antilepton fields, all of which operators conserve baryon number. There are a great number of such terms, but to violate lepton number conservation they must involve a product of two lepton fields or of their conjugates. The left-handed lepton doublets $(\ell^-_{Li}, \nu_i)$ and right-handed charged lepton singlets $\ell^-_{Ri}$ (with $i = e$, $\mu$, or $\tau$) have $U(1)$ quantum numbers $1/2$ and $+1$, respectively, while the scalar doublet (or doublets) $(\phi^+, \phi^0)$ have $U(1)$ quantum number $-1/2$, so we can construct $U(1)$-invariant interactions of dimensionality five out of two left-handed lepton doublets and two scalar doublets. With only a single type of scalar doublet, there is just one such term that satisfies $SU(2)$ and Lorentz invariance:[27a]

$$\sum_{ij} f_{ij} (\overline{\ell^c_{Li}} \phi^+ - \overline{\nu^c_i} \phi^0)(\ell_{Lj} \phi^+ - \nu_j \phi^0) , \qquad (21.3.54)$$

where $i$ and $j$ are lepton flavor indices, and $c$ denotes the charge conjugate field. At energies below the electroweak breaking scale, this yields an effective interaction

$$\sum_{ij} f_{ij} \, \overline{\nu^c_i} \nu_j \, \langle \phi^0 \rangle^2 \, . \qquad (21.3.55)$$

We expect $f_{ij}$ to be of order $1/M$, perhaps multiplied with small coupling constants, so this gives lepton number non-conserving neutrino masses at most of order[27b] $(300 \text{ GeV})^2/M$. We shall see in Section 21.5 that $M$

is expected to be of order $10^{15}$–$10^{18}$ GeV, so we would expect neutrino masses in the range $10^{-4}$–$10^{-1}$ eV, or less if suppressed by small coupling constants. Such masses are too small for direct measurement, but there is no reason for the neutrino mass matrix to be diagonal, so neutrino masses might be detected in oscillations of one neutrino type into another over long flight paths.

A similar analysis shows that there are interactions of dimensionality six that violate both baryon and lepton number conservation, involving three quark fields and one lepton field.[27c] Such interactions would have coupling constants of order $M^{-2}$, and would lead to processes like proton decay, with rates proportional to $M^{-4}$.

## 21.4 Dynamically Broken Local Symmetries*

Our discussion of spontaneously broken local symmetries has so far been entirely within the context of perturbation theory. To some extent, this limitation is inevitable. Whereas for spontaneously broken global symmetries it is possible to prove exact theorems about the existence and interactions of massless Goldstone bosons, the spontaneous breakdown of a local symmetry does not lead to any such precise consequences. Even the existence of massive vector bosons is not really a general theorem; for sufficiently strong gauge coupling these particles decay so rapidly that they lose their identity as distinct resonances of definite spin $j = 1$.

On the other hand, if the gauge couplings like $e$ or $g$ or $g'$ are sufficiently small then the theory with a spontaneously broken local symmetry must be very close to one with a spontaneously broken global symmetry, about which exact theorems can be proved. It is therefore possible to derive useful approximate results for such gauge theories, *even if the other non-gauge couplings are very strong*. One example is provided by the standard $SU(2) \times U(1)$ electroweak theory with a large scalar self-coupling $\lambda$ (and hence a large scalar mass; see Eq. (21.3.27)). A more intriguing possibility is that the breakdown of electroweak symmetry is due to strong forces associated with some new gauge group acting on a set of new fermions. We will here consider the results that can be obtained for all such theories, without reference to the specific mechanism for spontaneous symmetry breaking.[28]

We assume that in the limit of zero gauge couplings, our theory is invariant under some group $G$ of *global* symmetries, spontaneously broken

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.