SLAC-PUB-2544
June 1980
(T/E)

Grand Unified Theories and Proton Decay*

Paul Langacker

Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94305
and
University of Pennsylvania
Dept. of Physics, Philadelphia, PA 19104

(Submitted to Physics Reports)

*Work supported by the Department of Energy, contract numbers DE-AC03-76SF00515 and EY-76-C-02-3071.

ABSTRACT

Grand unified theories, in which the strong and electroweak interactions are embedded into an underlying theory with a single gauge coupling constant, are reviewed. A detailed description is given of many of the necessary background topics, including gauge theories, spontaneous symmetry breaking, the standard $SU_2 \times U_1$ electroweak model and its modifications and extensions, Majorana and Dirac neutrino masses, the induced cosmological term, CP violation, quantum chromodynamics and its symmetries, and dynamical symmetry breaking. The Georgi-Glashow SU_5 model is examined in detail. Models based on unitary, orthogonal, exceptional, and semi-simple groups and general constraints on model building are surveyed. Phenomenological aspects of grand unified theories are described, including the determination of the unification mass, the prediction of $\sin^2\theta_W$ in various models, existing and planned nucleon decay experiments, the predictions for the proton lifetime and branching ratios, general baryon number violating interactions, and the possible explanation of the matter-antimatter asymmetry of the universe. Other aspects of grand unified theories are discussed, including horizontal symmetries, neutrino and fermion masses, topless models, asymptotic freedom, implications for the neutral current, CP violation, superheavy magnetic monopoles, dynamical symmetry breaking, and the hierarchy problem.

Table of Contents

I. Introduction

1.1 Motivations and Introduction

One of the great triumphs of physics in recent years has been the development of the standard model of elementary particles and their interactions. The standard model combines the SU_3^c color gauge theory of the strong interactions (quantum chromodynamics) with the $SU_2 \times U_1$ model of weak and electromagnetic interactions. The standard model is compatible with all known facts of elementary particle physics and is believed to be free of mathematical inconsistencies. Moreover, the basic interactions are all described by gauge theories, which implies that the forms of the couplings of the vector bosons that mediate the interactions are determined by the underlying gauge symmetry.

Despite these attractive features, few physicists believe that the standard model is the ultimate theory of elementary particle interactions. The problem is not that there is anything wrong with the standard model, but rather that it is too complicated and arbitrary. In the standard model the strong, weak, and electromagnetic interactions are largely independent of each other, as is illustrated by the fact that the gauge group $G_s = SU_3^c \times SU_2 \times U_1$ is a direct product of three factors with different gauge coupling constants. Another problem is that the pattern of fermion representations is arbitrary and complicated. There is therefore no fundamental explanation for the repetition of fermion families or for the violation of parity in the weak interactions but not in the strong interactions. Yet another difficulty is that electric charges are not quantized. That is, there is no a prior reason for the quark and lepton charges to be related by simple factors of 3. Even given the groups, the representations, and the electric charge assignments, the standard model has many free parameters (19 or 26, depending on whether the neutrinos are assumed to be massless). This means that many observable quantities, such as the fermion masses, mixing angles, and CP violating phase, are completely arbitrary. Finally, the standard model does not incorporate gravity, nor does it offer an explanation for the empirical absence of a large cosmological term.

One way to constrain or determine some of the features that are arbitrary in the standard model is to consider models with more symmetry. Especially promising are grand unified theories (GUTS), in which the strong, weak, and electromagnetic interactions are embedded in a larger underlying gauge theory with a single gauge coupling constant g [1.1]. Grand unified theories typically have new symmetry generators which relate SU_3^c color quantum numbers and $SU_2 \times U_1$ flavor quantum numbers. Quarks, antiquarks, leptons, and antileptons are often placed together in irreducible representations of the underlying group G and are related by the new symmetries. Associated with the new symmetry generators are vector bosons which carry both flavor and color quantum numbers. In most grand unified theories these bosons, called X bosons, can mediate baryon number violating processes such as proton or bound neutron decay. A typical mechanism is for a quark to be transformed into a positron upon the emission of an X boson. Another quark then absorbs the X boson and changes into an antiquark. Such processes lead to decays such as $p \rightarrow e^+\pi^0$, with a nucleon lifetime proportional to M_X^4. The very stringent experimental limits on the proton lifetime ($\tau_p > 10^{30}$ yr) require that these new interactions must be extremely weak. One generally needs $M_X > 10^{14}$ GeV, more than twelve orders of magnitude larger than the masses of the bosons believed to be responsible for the weak interactions!

Fortunately, there is an independent way to predict M_X. In grand unified theories the observed differences between the strong, weak, and electromagnetic interactions, or between quarks, antiquarks, leptons, and antileptons, are due to the pattern of spontaneous symmetry breaking of the underlying gauge group G. If one could do experiments at momentum scales $Q^2 \gg M_X^2$, for which spontaneous symmetry breaking could be ignored, then all fermions would look very much alike and all interactions would be basically similar. In particular, the running coupling constants of the standard model interactions should all come together (up to normalization factors) at M_X. M_X can therefore be estimated from the observed low energy coupling constants and their theoretically known Q^2 dependence. Because of the large difference between the strong and weak couplings at low energy and their logarithmic Q^2 behavior, one finds very large values for M_X. In most models M_X is predicted to be in the range 10^{14-15} GeV! This is large enough to satisfy the existing limits on τ_p but small enough to suggest that proton decay may be observable in a new generation of experiments.

Grand unified theories have several appealing features. The basic idea of combining the strong and electroweak interactions into an underlying unified theory is very attractive and it makes the observed pattern of low energy interactions somewhat less arbitrary. Most grand unified theories explain the quantization of electric charge, and many of the models predict $\sin^2\theta_W \simeq 0.20$-$.21$, which is reasonably close to the experimental value ($0.229 \pm .014$) obtained from neutral current experiments. Most models (with the exception of the simplest SU_5 model) predict massive neutrinos, with masses typically in the range 10^{-5} - 10^{+2} eV. The associated neutrino oscillations could account for the missing solar neutrinos. Finally, some of the more complicated models predict flavor changing neutral currents (FCNC) at some level.

One of the most exciting implications of grand unified theories is that they may explain the excess of matter over antimatter in the universe (the baryon asymmetry). If baryon number were exactly conserved, the baryon excess would have to be postulated as an asymmetric initial condition on the big bang. (An alternative would be that baryons and antibaryons are separated over very long distance scales, but no plausible separation mechanism has been found). However, if grand unified theories are correct then the net baryon number of the universe could have been generated dynamically by baryon number violating interactions in the first instant after the big bang. This would have occurred for temperatures $T \lesssim M_X$, which corresponds to a time $t \gtrsim 10^{-35}$ sec after the big bang!

Grand unified theories also have their shortcomings. The models considered so far have shed little light on the pattern of fermion masses and mixings, or on the repetition of families. The appearance of two very different mass scales (M_W and M_X) is not a natural feature. It appears necessary to put this in by hand by fine tuning parameters, and it is not certain that this is even possible when radiative corrections are included (the hierarchy problem). Finally, gravity is not included in the usual models.

The unification of the strong and electroweak interactions apparently requires a bold (or perhaps foolhardy) extrapolation to energy scales many orders of magnitude beyond any that have been probed in the laboratory. (It is at least encouraging that the unification mass M_X turns out to be sufficiently small compared to the Planck mass $m_P \cong 1.22 \times 10^{19}$ GeV to justify, a posteriori, the neglect of quantum gravity.) It is very possible that some of the ideas of grand unification are wrong, at least in detail. Nevertheless, the features of grand unified models are so attractive and the consequences so dramatic that they deserve serious consideration.

1.2 Outline

This article describes most aspects of grand unification, with an emphasis on physical and phenomenological issues rather than group theory.

Chapter 2, which is a review in its own right, is a detailed description of the standard model and its modifications and extensions. The reader who is already familiar with the details of the standard model should skip most of this chapter and begin with Section 2.6. Sections 2.1 and 2.2 contain introductory comments and a summary of the elementary fields, respectively. Section 2.3 is an introduction to gauge theories. Global symmetries in field theory are described in 2.3.1, abelian and non-abelian local symmetries and the Feynman rules that they imply are considered in 2.3.2.

Spontaneous symmetry breaking via the Higgs mechanism is discussed in Section 2.3.3, along with ghosts, R_ξ gauges, and boson propagators. Section 2.4 deals with the electroweak interactions. QED is briefly summarized in 2.4.1. Relevant features of the weak interactions, including the formalism of P, C, and CP and early non-renormalizable theories, are considered in 2.4.2. The standard $SU_2 \times U_1$ electroweak theory of Glashow, Weinberg, Salam, and others is described in detail in Section 2.4.3. The history, structure, currents, Higgs potential, gauge bosons, fermion masses and mixings, phenomenology, and bounds on the Higgs mass are all discussed. Generalized electroweak theories and other theoretical issues are considered in 2.4.4. The topics discussed include the effects of additional Higgs and fermion representations, the induced cosmological term, anomaly constraints, Majorana and Dirac neutrino masses, larger weak groups, horizontal symmetries,and CP violation and the θ problem. Section 2.5 deals with the strong interactions. 2.5.1 is a descriptive introduction. Section 2.5.2 on quantum chromodynamics (QCD) describes the SU_3^c color group, asymptotic freedom, the strong coupling constant, and the symmetries of QCD, including chiral symmetry. The running coupling constants for arbitrary gauge groups are also considered in this section. Section 2.5.3 is devoted to dynamical symmetry breaking via technicolor (TC) and extended technicolor (ETC) interactions. The standard model is briefly summarized in 2.6.1, and its shortcomings and unanswered questions are described in Section 2.6.2.

The various grand unified models are discussed in detail in Chapter 3. After some general comments in 3.1 the formalism of the

SU_n groups is reviewed in 3.2. Section 3.3 is devoted to the simplest and most popular theory, the Georgi-Glashow SU_5 model. The basic structure of the model is described in Section 3.3.1; the gauge bosons, fermion representations, currents, proton decay mechanisms, spontaneous symmetry breaking pattern, Higgs potential, coupling constant predictions and fermion masses and mixings are all described. The successful and unsuccessful features of the model are summarized in 3.3.2, along with an indication of how these features are modified in other models.

Other grand unified models are described in Section 3.4. General constraints on model building are described in 3.4.1. This includes a classification of simple Lie groups, a listing of rank 4 groups with a single coupling constant, and discussions of complex vs. real representations; anomalies; color, charge, and weak embeddings; and C, P, and CP properties. SU_n, $n > 5$, models are briefly discussed in 3.4.2. SO_n models are considered in 3.4.3, with particular attention to the properties and symmetry breaking patterns of the SO_{10} model. The exceptional groups are described in 3.4.4, especially the phenomenology of the various E_6 models. Models based on semi-simple groups, such as the Pati-Salam models, are considered briefly in 3.4.5. These models are somewhat out of the mainstream, with some versions leading to integer charged quarks and/or low unification masses (10^{4-6} GeV).

The rest of the article deals with the phenomenology of grand unification. Proton decay is considered in Chapter 4. Section 4.1 contains general comments on baryon number, while 4.2 summarizes the existing limits and proposed experiments. The determination of the unification mass in the SU_5 model from the observed low energy coupling constants is described in 4.3.1. Included are the effects of Higgs particles, thresholds, renormalization schemes, the variation of the electromagnetic coupling, two loop effects, and uncertainties due to heavy bosons and fermions. The prediction of $\sin^2\theta_W$ and the consistency between α/α_s, $\sin^2\theta_W$, τ_p, and m_b/m_τ is discussed. The predictions for the nucleon lifetimes and branching ratios in SU_5 and related models are reviewed in 4.3.2. Included are the effective interaction, its symmetries and their consequences, selection rules, polarization predictions, anomalous dimensions, lifetime estimates, branching ratio estimates, and mixing effects. Baryon number violation is considered more generally in 4.4. Included are ways to avoid proton decay, B-L conserving and violating decays, and $\Delta B=2$ interactions which could lead to neutron-antineutron oscillations.

The baryon asymmetry of the universe is described in Chapter 5. The astrophysical and elementary particle components of the estimate of the baryon number to entropy ratio are surveyed, as are the possible washout of any initial asymmetry and the possibility of subsequent dilution effects.

Miscellaneous issues are described in Chapter 6. 6.1 deals with the fermion spectrum. An important technical ingredient, the survival hypothesis, is described in 6.1.1. Models which try to explain the repetition of fermion families by introducing horizontal interactions are described in 6.1.2. Neutrino masses are considered in 6.1.3 and the masses of other fermions in 6.1.4. Predictions for $\sin^2\theta_W$ in general models, including those with more than two mass scales, are outlined in 6.2. Topless models, asymptotic freedom, and modifications of the weak neutral current are described in Sections 6.3-6.5. CP violation is considered in 6.6. The θ problem, the probable restoration of spontaneously violated CP, and the associated difficulties with domains in the universe are described. The superheavy magnetic monopoles predicted by grand unified theories, the danger that they were produced too prolifically in the early universe, and possible solutions to this problem are described in 6.7. Attempts to embed technicolor into grand unified theories are described in 6.8. The hierarchy problem is surveyed in 6.9.

The article concludes in Chapter 7 with a summary of outstanding problems and a brief mention of other directions not covered in this article.

1.3 Summary of Notation

I have attempted to use consistent notation and conventions throughout this article, although the alert reader will find occasional inconsistencies. My metric and Dirac matrix conventions are those of Bjorken and Drell [1.2], except for the sign of γ_5.

For the many spaces of indices, I have used the following conventions: (μ, ν, σ, τ) are Lorentz indices; (i, j, k, = 1,...,N) label group generators; (a, b, c) run over the elements of a representation, while (α, β, γ) and (r, s, t) are used for the special cases of color and flavor indices, respectively; (m, n, p, q) are horizontal indices, labeling repeated fermion or Higgs representations. The summation convention applies to all repeated indices in all spaces.

Antiparticles are usually represented by a superscript c rather than an overbar (e.g., u^c is the anti u quark).

Operators are represented by capital letters (T^i, Q, Y), their eigenvalues by lower case letters (t^i, q, y), and their matrix representations by (L^i, L_Q, L_Y).

Yukawa couplings are generally denoted by Γ, γ, or η.

Gauge boson and Higgs boson masses are represented by M and μ, respectively (except in Chapter 5). Fermion masses and mass matrices are usually represented by m and M, respectively (e.g. m_u is the u quark mass and M^u is the charge 2/3 mass matrix).

Certain acronyms used frequently are FCNC = flavor changing neutral current, VEV = vacuum expectation value, SSB = spontaneous symmetry breaking, IRREP = irreducible representation, (E)TC = (extended) technicolor, GUT = grand unified theory.

In Feynman diagrams, spin-0, -½, -1, and ghost fields are represented by dashed, solid, wavy, and curly lines, respectively.

I have typically translated the quantitative results of other authors into my conventions or into common formats for comparison and have occasionally corrected trivial errors without drawing attention to them.

2. THE STANDARD MODEL AND ITS LIMITATIONS

2.1 Introductory Remarks

The last decade or so has witnessed a tremendous advance in our understanding of elementary particles and their interactions. In the late 1960's the known elementary particles were the leptons (the electron, the muon, and their neutrinos), the photon, and a great variety of hadrons, or strongly interacting particles. (The quark model of hadrons was well known but far from universally accepted.) Of the four known interactions, there existed a completely satisfactory theory of only one, the electromagnetic. A great wealth of experimental data on the charged-current weak interactions was phenomenologically described by the Cabibbo theory, which was the latest form of the current-current theory of weak interactions originally proposed by Fermi. However, it was known that the Cabibbo theory could only be valid at low energies: amplitudes would grow with energy and eventually violate the unitarity limit. Unitarity could not be restored simply by including higher order diagrams, because the four-Fermion interaction theory was non-renormalizaable (i.e., higher order Feynman diagrams involved severe ultraviolet divergences that could not be absorbed into the renormalization of a finite number of masses and coupling constants). Even the hypothesis that the weak interactions were really mediated by massive intermediate vector bosons did not cure the unitarity and nonrenormalizability diseases. The Weinberg-Salam theory of leptons was published in 1967 (but largely ignored until 1971), but it was not known at that time that the theory was renormalizable or how it could be extended to hadrons.

The situation was even worse for the strong interactions. Many systematic features of the hadronic spectrum and interactions were known and many theoretical models existed. States were classified according to SU_3 and higher symmetries such as SU_6 (which related the spin and internal degrees of freedom), as well as according to the nonrelativistic quark model. Regge and dual models further related states of different spin. Many dynamical models, such as current algebra, PCAC, vector meson dominance, Regge theory, dual resonance models, bootstrap theory, finite energy sum rules, one boson exchange models, and the parton model described, with more or less success, some limited aspects of hadronic physics, but none (with the possible exception of the bootstrap) could be considered a complete or fundamental theory of all aspects of the strong interactions.

The situation at present is completely different. There now exists a theory of the strong, weak, and electromagnetic interactions--the standard model--which most physicists believe to be a substantially correct and perhaps fundamental description of all interactions except gravity. In the standard model, the strong, weak, and electromagnetic interactions are all gauge interactions, which means in part that they are mediated by the exchange of vector bosons. The standard model has no known inconsistencies or mathematical diseases; it really is a candidate for a fundamental theory of physics (one must remain skeptical, however, of any theory which fails to incorporate gravity). Furthermore, the standard model successfully describes an enormous number of quantitative and qualitative features of elementary particle physics. There are no known experimental facts that conflict with the standard model, although many aspects of the model, especially in the strong interaction sector, are beyond our ability to calculate at present.

Assuming that the standard model correctly describes all "low-energy" physics ("low energy" includes all energies that have been explored to date) one must still ask whether the standard model is the entire story or whether it is just the low energy limit of a bigger theory. The difficulty with the first point of view is that the standard model leaves an uncomfortable number of important questions unanswered. There is no explanation or prediction of the number of elementary fields, for example. Furthermore, the minimal version of the model contains 19 arbitrary parameters (26 if one allows the neutrinos to have masses). Finally, the strong, weak, and electromagnetic interactions are basically unrelated and independent of each other in the standard model. It is, of course, possible that all of these quantities really are arbitrary or that their origin is simply too hard for us to understand. Nevertheless, it was the attempt to understand some of these questions that has led to the development of grand unified theories. The idea is that the observed interactions are merely the low energy manifestation of an underlying unified theory. This underlying theory possesses additional structure that may constrain some of the quantities that appear arbitrary at the level of the standard model.

Such underlying unified theories are the main topic of this report. I must begin, however, by describing the standard model in detail. In Section 2-2 of this chapter I will describe the elementary fields or building blocks of the standard model. Section 2-3 contains a review of the ideas and formalism of gauge theories. In Section 2-4 the principal features of the weak and electromagnetic interactions and of the Glashow-Weinberg-Salam (GWS) model are described. Various extensions and modifications of the GWS model (especially those which are relevant to grand unification) are also discussed.

Section 2-5 is a description of the strong interactions and of Quantum Chromodynamics (QCD), as well as a brief discussion of dynamical symmetry breaking. Finally, in Section 2.6 the standard model is described; its weaknesses and the motivations for grand unification are further discussed.

2.2 The Elementary Fields

The fundamental quantum fields which appear in the standard model are of four types:

gauge bosons (spin-1): γ, $W^{\pm}$, Z, G^{α}_{β}

leptons (spin-1/2): e^{-}, μ^{-}, τ^{-}, ν_e, ν_μ, ν_τ

quarks (spin-1/2): q^{α}_{r} $\left\{\begin{array}{l}\alpha = R,\ G,\ B = \text{color} \\ r = u,\ d,\ c,\ s,\ t,\ b = \text{flavor}\end{array}\right.$

Higgs bosons (spin 0): ϕ^{+}, ϕ^{o}

The gauge or vector bosons mediate the various interactions. The photon γ is responsible for electromagnetism, while the intermediate vector bosons $W^{\pm}$ and Z mediate charged and neutral current weak interactions, respectively. The strong interactions are ultimately due to the exchange of eight gluons G^{α}_{β}, where the color indices α and β assume the values 1, 2, and 3 or red (R), green (G) and blue (B). There are only eight gluon fields because the G are constrained by $\sum_{\alpha} G^{\alpha}_{\alpha} = 0$.

The basic Fermions are the leptons, or non-strongly interacting particles, and the quarks, which are believed to be the constituents of the hadrons (the strongly interacting particles). The quarks carry two types of internal quantum number or index. The flavor quantum number r represents all of the internal quantum numbers conserved by the strong interactions. The u (up) and d (down) quarks carry $\pm$ 1/2 unit of I^3, the third component of isospin, and are the constituents of the proton, neutron, pion, etc. The heavier flavors of quarks include the s (strange), c (charm), b (bottom), and t (top) quark. Baryons are believed to be bound states of three quarks (e.g., p = uud, n = udd, Σ^+ = uus) and mesons involve a quark-antiquark pair ($\pi^+ = u\bar{d}$, $K^+ = u\bar{s}$, $D^+ = c\bar{d}$, $J/\psi = c\bar{c}$, $\Upsilon = b\bar{b}$, etc.). The u,c, and t quarks have electric charge 2/3, while the d, s, and b quarks have charge -1/3.

Each flavor of quark comes in three color states. The wave functions of physical hadrons are arranged so that the hadrons are neutral with respect to color. For example, ignoring the space and spin degrees of freedom, the proton and π^+ wave functions are

$$\begin{aligned} p = \frac{1}{\sqrt{6}} (&u^R u^G d^B + u^G u^B d^R + u^B u^R d^G \\ &- u^R u^B d^G - u^B u^G d^R - u^G u^R d^B) \\ \pi^+ = \frac{1}{\sqrt{3}} &(u^R \bar{d}^R + u^G \bar{d}^G + u^B \bar{d}^B) \end{aligned} \tag{2.1}$$

Note that the baryon wave functions are antisymmetric with respect to color. This was one of the original motivations for introducing the color quantum number. Quark model classifications of the baryon spectrum require that the wave functions be totally symmetric in the spin, flavor, and spatial variables (assuming that the ground states have zero orbital angular momentum). Hence, Fermi statistics suggests the existence of an additional quantum number in order to generate an antisymmetric wave function.

Of the particles described so far, only the photon and the leptons have been directly observed. (The ν_τ has not yet been unambiguously observed.) The $W^\pm$ and Z will hopefully be discovered at the next generation of accelerators and storage rings. The gluons and quarks, which carry color, are believed to not exist as isolated states. However, indirect evidence exists for all of the quark flavors except the top, which (at the time of this writing) is only a theoretical speculation.

The hypothetical and mysterious spin-0 Higgs fields are introduced into the theory to give masses to the $W^{\pm}$ and Z bosons. They will be discussed in Section 2.4.

2.3 Gauge Theories

In this section I give a review of the formalism of gauge theories. For a more detailed introduction the reader is referred to the many excellent reviews available, such as those of Abers and Lee [2.1], Bég and Sirlin [2.2], Bernstein [2.3], Iliopoulos [2.3], Marciano and Pagels [2.5], Taylor [2.6], Weinberg [2.7,2.8], and Fradkin and Tyutin [2.8a].

2.3.1 Global Symmetries [2.9]

Consider a Lie group G with generators T^i, $i=1,2,3...N$. The structure constants c_{ijk} of G are defined by

$$[T^i, T^j] = i\, c_{ijk}\, T^k , \tag{2.2}$$

where a summation on k from 1 to N is implied. If the generators are chosen to be Hermitian then the structure constants are real. A set of matrices L^i which satisfy the same commutation relations

$$[L^i, L^j] = i\, c_{ijk}\, L^k \tag{2.3}$$

as the generators form a representation of G. It will be convenient to define and normalize the generators so that

$$\mathrm{Tr}(L^i\, L^j) = T(L)\,\delta_{ij} \quad , \tag{2.4}$$

with $T(L) = \frac{1}{2}$ for the lowest dimensional representation (this is always possible for the semisimple and U_1 Lie algebras with which we will be concerned [2.10]). It is then easy to see that c_{ijk} is totally antisymmetric in all three indices.

An arbitrary element of G is specified by N continuous real parameters $\beta^1, \ldots, \beta^N$. Define the N component vector $\vec{\beta} \equiv (\beta^1, \ldots, \beta^N)$. Then an arbitrary element of G is

$$U_G(\vec{\beta}) = \exp(i \sum_{i=1}^{N} \beta^i\, T^i) = \exp(i\, \vec{\beta} \cdot \vec{T}). \tag{2.5}$$

Now consider a set of n fields $\Phi_a(x)$, $a=1,\ldots,n$, (which may be either fermion or boson fields) which transform according to the n x n dimensional representation matrices L^i. This means that

$$[T^i, \Phi_a(x)] = -\, L^i_{ab}\, \Phi_b(x) \ . \tag{2.6}$$

Under a group transformation

$$\begin{aligned} \Phi_a(x) \to \Phi'_a(x) &\equiv U_G(\vec{\beta})\, \Phi_a(x)\, U_G(\vec{\beta})^{-1} \\ &= \Phi_a + i[\vec{\beta}\cdot\vec{T}, \Phi_a] + \frac{i^2}{2!} [\vec{\beta}\cdot\vec{T}, [\vec{\beta}\cdot\vec{T}, \Phi_a]] \\ &\quad + \ldots \\ &\quad + \frac{i^k}{k!} \underbrace{[\vec{\beta}\cdot\vec{T}, [\ldots [\vec{\beta}\cdot\vec{T}, \Phi_a]]}_{k} + \ldots \end{aligned} \tag{2.7}$$

Hence, under an infinitesimal transformation,

$$\Phi_a(x) \rightarrow \Phi_a - i(\vec{\beta}\cdot\vec{L})_{ab}\,\Phi_b + O(\beta^2), \tag{2.8}$$

while for a finite transformation

$$\Phi_a(x) \rightarrow (e^{-i\vec{\beta}\cdot\vec{L}})_{ab}\,\Phi_b(x) \equiv U(\vec{\beta})_{ab}\,\Phi_b(x) \tag{2.9}$$

In (2.6) Φ_a includes all of the boson or fermion fields in the theory. If the Φ_a are complex it is often (but not always) the case that the representation matrices are reducible into two n/2 x n/2 dimensional sectors, with n/2 fields transforming according to

$$[T^i, \Phi_a] = -\,\ell^i_{ab}\,\Phi_b\;, \tag{2.10}$$

$(a,b) \in 1,\ldots,n/2$, and their Hermitian conjugates $\Phi^\dagger_a$ transforming as

$$[T^i, \Phi^\dagger_a] = +\,\ell^{i*}_{ab}\,\Phi^\dagger_b \;=\; -(-\ell^{iT}_{ab})\,\Phi^\dagger_b\;, \tag{2.11}$$

where I have taken T^i (and ℓ^i) to be Hermitian and ℓ^{iT} is the transpose of ℓ^i. Hence, $\Phi^\dagger_a$ transforms according to $-\ell^{iT}$. Of course, ℓ^i and $-\ell^{iT}$ may be equal or equivalent (related by a similarity transformation). Such representations are called real.

If the Lagrangian (and therefore the equations of motion) of a theory is left invariant under the replacement $\Phi_a \rightarrow \Phi'_a$ then according to the Noether theorem [2.11] there exist N conserved currents J^i_μ associated with the generators. They are given by

$$J^i_\mu = -i\,\frac{\delta\mathscr{L}}{\delta\,\partial^\mu\,\Phi_a}\,L^i_{ab}\,\Phi_b\;. \tag{2.12}$$

The associated charges

$$T^i \equiv \int d^3 \underset{\sim}{x}\; J_o^i(x) \tag{2.13}$$

are the generators of the group.

For example, suppose $\mathscr{L}$ involves n Fermion fields ψ_a, a=1,...,n and m complex scalar fields ϕ_b, b=1,...,m. (Scalar simply means spin 0; the ϕ_b can have either positive or negative intrinsic parity or even no definite parity.) Then

$$\mathscr{L} = \sum_{a=1}^{n} \bar{\psi}_a\, i\, \partial\!\!\!/\, \psi_a + \sum_{b=1}^{m} (\partial_\mu \phi_b)^\dagger \partial^\mu \phi_b \tag{2.14}$$

$$+ \text{ non-derivative terms.}$$

Then, if ℓ_ψ^i and ℓ_ϕ^i are the representation matrices for ψ and ϕ respectively (I have assumed that neither ϕ_b and $\phi_b^\dagger$ nor the ψ_a and $\psi_a^\dagger$ are rotated into each other by the symmetry transformations.) then the Noether currents are

$$J_\mu^i = \sum_{a,b=1}^{n} \bar{\psi}_a\, \gamma_\mu (\ell_\psi^i)_{ab}\, \psi_b + i \sum_{a,b=1}^{m} \phi_a^\dagger (\ell_\phi^i)_{ab} \overleftrightarrow{\partial}_\mu \phi_b \,, \tag{2.15}$$

where

$$f \overleftrightarrow{\partial}_\mu g \equiv f(\partial_\mu g) - (\partial_\mu f) g \,. \tag{2.16}$$

It is often convenient to work in terms of Hermitian scalar fields $\phi_b = \phi_b^\dagger$ (a complex field Φ can always be written as $\Phi = \frac{1}{\sqrt{2}}(\phi_R + i\,\phi_I)$, where ϕ_R and ϕ_I are Hermitian). From (2.11) one then has that the representation matrices L^i are antisymmetric $L^i = -L^{iT}$. For Hermitian fields, the kinetic energies term

$$\mathscr{L}_{kin} = \sum_{b=1}^{n} \frac{1}{2} (\partial_\mu \phi_b)(\partial^\mu \phi_b) \tag{2.17}$$

implies the Noether current

$$J_\mu^i = -i \sum_{a,b=1}^{n} (\partial_\mu \phi_a) L_{ab}^i \phi_b \tag{2.18}$$

For a theory involving fermions, there is no need for the left and right helicity projections of the field

$$\psi_{L,R} \equiv P_{L,R}\, \psi \equiv \frac{1 \pm \gamma_5}{2} \psi \tag{2.19}$$

to transform according to the same representation. That is, for

$$[T^i, \psi_{aL,R}] = -(\ell^i_{L,R})_{ab}\, \psi_{bL,R} , \tag{2.20}$$

the Noether current is

$$\begin{aligned} J_\mu^i &= \bar{\psi}_{aL}\, \gamma_\mu (\ell_L^i)_{ab} \psi_{bL} + \bar{\psi}_{aR} \gamma_\mu (\ell_R^i)_{ab}\, \psi_{bR} \\ &= \bar{\psi}_a\, \gamma_\mu (\ell_L^i\, P_L + \ell_R^i\, P_R)_{ab}\, \psi_b \end{aligned} \tag{2.21}$$

However, fermion mass terms, which are of the form

$$\bar{\psi}_{aL}\, m_{ab}\, \psi_{bR} + \text{H.C} \tag{2.22}$$

are not invariant if ψ_{aL} and ψ_{bR} transform according to different irreducible representations. Such transformations are referred to as chiral transformations.

If G is an exact symmetry of the Lagrangian then the Noether currents J_μ^i are conserved and the charges T^i are constants. Hence, T^i can represent a conserved quantum number if the symmetry is not spontaneously broken (i.e., broken in the vacuum; see Section 2.3.3). In fact, only those generators that can be simultaneously diagonalized will correspond to quantum numbers. The maximal number of generators that commute with each other (and therefore can be simultaneously diagonalized) is called the rank of G. The rank is therefore equal to the number of conserved quantum numbers associated with G.

2.3.2 Local Symmetries

The symmetry transformations described in the last section are known as global symmetries because the fields transform in the same way at every point in space and time. That is

$$\Phi_a(x) \rightarrow (e^{-i\vec{\beta}\cdot\vec{L}})_{ab}\,\Phi_b(x) \ , \tag{2.23}$$

where $U(\vec{\beta}) = \exp(-i\vec{\beta}\cdot\vec{L})$ is the same for all x.

If the symmetry is extended to allow independent transformations at different space-time points, the symmetry is known as a local or gauge symmetry. [2.12] Under a gauge transformation

$$\begin{aligned}\Phi_a(x) &\rightarrow (e^{-i\vec{\beta}(x)\cdot\vec{L}})_{ab}\,\Phi_b(x) \\ &= U(\vec{\beta}(x))_{ab}\,\Phi_b(x) \ ,\end{aligned} \tag{2.24}$$

where $\vec{\beta}(x)$ is now an arbitrary differentiable function of x.

The requirement that a Lagrangian be invariant under a local symmetry is much more stringent than the requirement of global invariance. In fact, the existence of a local symmetry implies the existence of (apparently) massless vector bosons (or gauge bosons), one for each generator of the local symmetry group. Furthermore, the structure of the interactions of the gauge bosons with each other and with other fields are prescribed by the gauge invariance.

Abelian Local Symmetries

Consider first the Lagrangian for a free Fermion field

$$\mathscr{L} = \bar{\psi}(i\not{\partial} - m)\psi \ . \tag{2.25}$$

$\mathscr{L}$ is invariant under the global phase transformations

$$\psi(x) \rightarrow e^{-i\beta}\ \psi(x)\ , \tag{2.26}$$

where β is an arbitrary real number. These transformation form the abelian (commutative) group U_1 of unitary 1x1 matrices or phase factors. The associated Noether current is

$$J_\mu = \bar{\psi}\ \gamma_\mu\ \psi \tag{2.27}$$

and the conserved quantum number is particle number.

$\mathscr{L}$ as written is not invariant under the local U_1 transformation

$$\psi(x) \rightarrow e^{-i\beta(x)}\ \psi(x) \tag{2.28}$$

because of the derivative. Rather

$$\mathscr{L} \rightarrow \mathscr{L} + \bar{\psi}(\not{\partial}\beta)\psi \ . \tag{2.29}$$

In order to have an invariant Lagrangian it is necessary to replace ∂_μ by the gauge covariant derivative

$$D_\mu \equiv \partial_\mu - ig\ A_\mu(x)\ , \tag{2.30}$$

where $A_\mu(x)$ is a spin-1 field which transforms as

$$A_\mu(x) \rightarrow A_\mu(x) - \frac{1}{g}\,\partial_\mu\ \beta(x) \tag{2.31}$$

under the gauge transformation. The gauge coupling constant g is arbitrary.

The derivatives in (2.29) and (2.31) cancel so that

$$D_\mu \psi(x) \rightarrow e^{-i\beta(x)} D_\mu \psi(x) \ . \tag{2.32}$$

One must also add a gauge invariant kinetic energy term

$$\mathscr{L}_{kin} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{2.33}$$

to $\mathscr{L}$, where the gauge invariant field tensor is

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu \ . \tag{2.34}$$

Note, however, that a vector meson mass term

$$\mathscr{L}_{mass} = \frac{M^2}{2} A_\mu A^\mu \tag{2.35}$$

is forbidden by the requirement of gauge invariance. Hence, the final locally invariant Lagrangian is

$$\begin{aligned} \mathscr{L} &= -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i \not{D} - m)\psi \\ &= -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(i\not{\partial} + g\not{A} - m)\psi \ . \end{aligned} \tag{2.36}$$

We see that gauge invariance requires the existence of a massless vector field (or gauge field) A_μ. The form of the interaction of A_μ with the fermion field is also determined: A_μ couples to the Noether current $\bar{\psi} \gamma_\mu \psi$ with coupling constant g, as shown in Fig. 2.1. The massless boson of course generates a long range force.

Similarly, the Lagrangian for a free complex scalar field ϕ is

$$\mathscr{L} = (\partial^\mu \phi)^\dagger \partial_\mu\phi - \mu^2 \phi^\dagger \phi \ . \tag{2.37}$$

(For the U_1 group it is most convenient to use a complex basis). $\mathcal{L}$ is invariant under $\phi(x) \to \exp(-i\beta)\,\phi(x)$, with a Noether current

$$J_\mu = i\,\phi^\dagger \overleftrightarrow{\partial}_\mu \phi\ . \tag{2.38}$$

To obtain a gauge invariant Lagrangian one must again replace ∂_μ by the covariant derivative, so that

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi \\ &= -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + [(\partial^\mu + ig\,A^\mu)\phi^\dagger](\partial_\mu - igA_\mu)\phi - \mu^2\phi^\dagger\phi \\ &= -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \partial^\mu \phi^\dagger \partial_\mu \phi + ig(\phi^\dagger \overleftrightarrow{\partial}_\mu \phi)A^\mu + g^2\phi^\dagger\phi\, A^2 \\ &\qquad -\mu^2\phi^\dagger\phi\end{aligned} \tag{2.39}$$

The vertices for the emission or absorption of one or two gauge bosons are shown in Fig. 2.1.

For a theory with several complex scalar and fermion fields with charges q_a (in units of g), then invariance under

$$\begin{aligned}\psi_a(x) &\to e^{-iq_a\beta(x)}\,\psi_a(x) \\ \phi_b(x) &\to e^{-iq_b\beta(x)}\,\phi_b(x)\end{aligned} \tag{2.40}$$

requires the gauge invariant kinetic energy terms

$$\begin{aligned}\mathcal{L}_{kin} &= -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + \sum \bar{\psi}_a(i\,\not{\partial} + q_a g\,\not{A})\psi_a \\ &\quad + \sum_b [(\partial^\mu + i\,q_b\,g\,A^\mu)\phi_b^\dagger][\partial_\mu - i\,q_b\,g\,A_\mu]\phi_b\end{aligned} \tag{2.41}$$

To this one may add Yukawa terms

$$\mathscr{L}_{Yuk} = \bar{\psi}_a \, \Gamma^c_{ab} \, \psi_b \, \phi_c + \text{H.C.} \tag{2.42}$$

and fermion and scalar mass terms

$$\mathscr{L}_{mass} = - \bar{\psi}_a \, m_{ab} \, \psi_b - \phi^\dagger_c \, \mu^2_{cd} \, \phi_d \ , \tag{2.43}$$

provided they are invariant under (2.40). The matrices Γ^c and m may both involve γ_5's. That is,

$$\begin{aligned} \Gamma^c &= \Gamma^c_L \frac{1+\gamma_5}{2} + \Gamma^c_R \frac{1-\gamma_5}{2} \\ &= \Gamma^c_L \, P_L + \Gamma^c_R \, P_R \end{aligned} \tag{2.44}$$

and

$$m = m_L \, P_L + m_R \, P_R \ , \tag{2.45}$$

with $m_R = m^\dagger_L$.

One can also consider chiral U_1 gauge symmetries in which ψ_L and ψ_R transform differently (chiral transformations). If

$$\begin{aligned} \psi_L &\to e^{-i \, q_L \beta(x)} \, \psi_L \\ \psi_R &\to e^{-i \, q_R \beta(x)} \, \psi_R \end{aligned} \tag{2.46}$$

then the gauge invariant kinetic energy term for the ψ field is

$$\begin{aligned} \mathscr{L}_{kin} &= \bar{\psi}_L \, i \, \not{D}^L \, \psi_L + \bar{\psi}_R \, i \, \not{D}^R \, \psi_R \\ &= \bar{\psi}_L \, (i \, \not{\partial} + q_L \, g \, \not{A}) \, \psi_L + \bar{\psi}_R (i \, \not{\partial} + q_R \, g \, \not{A}) \psi_R \\ &= \bar{\psi} [i \, \not{\partial} + g \, \not{A} \, (g_L \, P_L + g_R \, P_R)] \psi \end{aligned} \tag{2.47}$$

Non-Abelian Local Symmetries

Let us now consider local symmetries based on non-abelian groups. The first example of such a theory was given by Yang and Mills in 1954 [2.13-14, 2.10], but the following discussion will parallel more closely the excellent treatments by Weinberg [2.7-8], Abers and Lee [2.1], and Bég and Sirlin [2.2].

Suppose the kinetic energy, mass, and Yukawa terms in a Lagrangian involving scalar and fermion fields are invariant with respect to a global symmetry group G. Let L_ψ and L_ϕ be the representation matrices for the fermions and scalars, so that

$$\psi_a \to (e^{-i\vec{\beta}\cdot\vec{L}_\psi})_{ab}\ \psi_b$$

$$= U_\psi(\vec{\beta})_{ab}\ \psi_b \tag{2.48}$$

$$\phi_a \to (e^{-i\vec{\beta}\cdot\vec{L}_\phi})_{ab}\ \ \phi_b = U_\phi(\vec{\beta})_{ab}\ \phi_b .$$

For a chiral transformation one can write

$$L^i_\psi = L^i_L\ P_L + L^i_R\ P_R\ . \tag{2.49}$$

It is convenient to write

$$\Phi_a \to U_{ab}(\vec{\beta})\ \Phi_b \tag{2.50}$$

or just

$$\Phi \to U(\vec{\beta})\ \ \Phi, \tag{2.51}$$

where Φ_a can represent either ψ_a or ϕ_a and Φ is a vector with the Φ_a as its components.

The mass and Yukawa terms will still be invariant under the local symmetry

$$\Phi(x) \to U(\vec{\beta}(x))\ \Phi(x) = e^{-i\vec{\beta}(x)\cdot\vec{L}}\ \Phi(x), \tag{2.52}$$

but the kinetic energy terms will not be invariant because of the presence of derivatives, which transform as

$$\partial_\mu \Phi \to U\partial_\mu \Phi + (\partial_\mu U)\ \Phi\ . \tag{2.53}$$

The Lagrangian can be made gauge invariant by replacing each derivative by a covariant derivative

$$D_\mu \equiv \partial_\mu - ig\ \vec{A}_\mu \cdot \vec{L}\ , \tag{2.54}$$

where A^i, $i=1,\ldots,N$ are N vector gauge fields, and the L^i are the representation matrices for the fields on which D acts. D is actually a matrix in the space of group indices:

$$\begin{aligned}(D_\mu\ \Phi)_a &= (D_\mu)_{ab}\ \Phi_b \\ &= (\partial_\mu\ \delta_{ab} - ig\ \vec{A}_\mu \cdot \vec{L}_{ab})\ \Phi_b\ .\end{aligned} \tag{2.55}$$

Under a gauge transformation the A^i transform in such a way as to cancel the ∂_μ U piece in the transformation of $\partial_\mu \Phi$

$$\begin{aligned}\vec{A}_\mu \cdot \vec{L} &\to \vec{A}'_\mu \cdot \vec{L} \\ &\equiv U\ \vec{A}_\mu \cdot \vec{L}\ U^{-1} - \frac{i}{g}\ (\partial_\mu\ U)\ U^{-1}\ .\end{aligned} \tag{2.56}$$

Then,

$$\begin{aligned}D_\mu\Phi &= (\partial_\mu - ig\ \vec{A}_\mu \cdot \vec{L})\ \Phi \\ &\to U[\partial_\mu + U^{-1}\ \partial_\mu U - ig\ \vec{A}_\mu \cdot \vec{L} - U^{-1}\ \partial_\mu U)\ \Phi \\ &= U\ D_\mu\ \Phi\ ,\end{aligned} \tag{2.57}$$

so that the kinetic energy terms for fermion and complex scalar fields

$$\mathscr{L}_{kin} = \bar{\psi}\, i\, D_{\psi}\, \psi + (D^{\mu}_{\phi}\ \phi)^{\dagger}\ D_{\phi\mu}\ \phi \tag{2.58}$$

are gauge invariant. The gauge invariance prescribes the form of the couplings between the A^i_μ and other fields. The three and four point vertices in (2.58), which are often off-diagonal, are shown in Fig. 2.2. I will henceforth simplify the notation by dropping the subscripts ψ or ϕ on $\vec{L}$, U, and D_μ.

The gauge invariant kinetic energy terms for Hermitian scalar fields are

$$\begin{aligned} \mathscr{L}_{kin} &= \frac{1}{2}\,(D^{\mu}\ \phi)\ D_{\mu}\ \phi \\ &= \frac{1}{2}\,\phi^{T}(\overleftarrow{\partial}^{\mu} + ig\ \vec{A}^{\mu}\cdot\vec{L})(\partial_{\mu} - ig\ \vec{A}_{\mu}\cdot\vec{L})\ \phi\ , \end{aligned} \tag{2.59}$$

where the antisymmetric property $L^i = -L^{iT}$ has been used. The vertices are shown in Fig. 2.2.

It might appear from (2.56) that the transformation of A^i_μ depends on which representation matrices L^i are used. This in fact is not the case, as can be seen by going to the infinitesimal form of (2.56)

$$\vec{A}_{\mu}\cdot\vec{L} \underset{\vec{\beta}\to 0}{\to} \vec{A}_{\mu}\cdot\vec{L} + \beta^{i}A^{j}_{\mu}\ L^{k}\ c_{ijk} - \frac{1}{g}\,\partial_{\mu}\ \vec{\beta}\cdot\vec{L}\ ; \tag{2.60}$$

projecting out L^i,

$$A^{i}_{\mu} \to A^{i}_{\mu} + c_{ijk}\ \beta^{j}\ A^{k}_{\mu} - \frac{1}{g}\,\partial_{\mu}\ \beta^{k}\ , \tag{2.61}$$

independent of the representation. (2.61) differs from the abelian case by the second term, which shows that the A^i_μ transform non-trivially even under global transformations according to the adjoint or regular representation of G, defined by

$$(L^i_{adj})_{jk} = -i\, c_{ijk} \tag{2.62}$$

For the N gauge fields one must introduce the kinetic energy term

$$\mathscr{L}_{kin} = -\frac{1}{4} F^i_{\mu\nu} F^{i\mu\nu} , \tag{2.63}$$

where

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g\, c_{ijk} A^j_\mu A^k_\nu \quad . \tag{2.64}$$

$\mathscr{L}_{kin}$ is gauge invariant provided that the c_{ijk} are totally antisymmetric, which is guaranteed by the convention (2.4). Abers and Lee [2.1], for example, show that under an infinitesimal transformation

$$F^i_{\mu\nu} \to F^i_{\mu\nu} + c_{ijk}\, \beta^j F^k_{\mu\nu} \quad , \tag{2.65}$$

from which the invariance follows. It is sometimes useful to consider the matrix

$$\vec{F}_{\mu\nu} \cdot \vec{L} \equiv \partial_\mu \vec{A}_\nu \cdot \vec{L} - \partial_\nu \vec{A}_\mu \cdot \vec{L} - ig[\vec{A}_\mu \cdot \vec{L}, \vec{A}_\nu \cdot \vec{L}] \; . \tag{2.66}$$

It is easy to show that

$$\vec{F}_{\mu\nu}(x) \cdot \vec{L} \to U(x)\, \vec{F}_{\mu\nu} \cdot \vec{L}\, U^{-1}(x) \; , \tag{2.67}$$

so that

$$\begin{aligned} \mathscr{L}_{kin} &= -\frac{1}{4} F^i_{\mu\nu} F^{i\mu\nu} \\ &= -\frac{1}{4T(L)} \mathrm{Tr}(\vec{F}_{\mu\nu} \cdot \vec{L})^2 \end{aligned} \tag{2.68}$$

is gauge invariant. The last term in F^i implies the existence of self-interactions between the gauge fields. The three and four point vertices are shown in Fig. 2.2 .

The gauge boson mass terms

$$\mathscr{L}_{mass} = \frac{1}{2} M_{ij}^2 \; A^{i\mu} \; A_\mu^j \qquad (2.69)$$

are not gauge invariant, so the N gauge bosons appear to be massless.

In summary then, a Lagrangian with a global symmetry group with N generators can be made locally invariant by introducing N (apparently) massless vector fields which have specified couplings to other fields and which transform according to the adjoint representation under global transformations. The non-abelian case therefore differs from the abelian in that the gauge fields themselves carry the "charges" associated with the generators of G. This leads to the existence of off-diagonal vertices in which a fermion or scalar field Φ_a absorbs or emits a gauge boson and turns into a different field Φ_b. Similarly, there are elementary self-interactions between the gauge bosons in the non-abelian case. Other aspects of gauge theories will be discussed later. In particular, a brief discussion of propagators and useful special gauges will be given in the next section, while running coupling constants will be described in Section 2.5. Before proceeding it should be remarked that if G can be written as the direct product of two or more smaller groups, $G = G_1 \times G_2 \times \ldots$, then one can have different coupling constants $g_1, g_2, \ldots$ for the interactions of the gauge bosons associated with each of the factors.

Applications

Gauge theories are very attractive in that the structure of the gauge interactions is dictated by the gauge invariance. Furthermore, they are believed to be the only field theories for vector mesons that are renormalizable [2.15-17], which means that all of the ultraviolet divergences in higher order diagrams can be removed from the theory by the redefinition of a finite number of masses and coupling constants. (Renormalizability also requires the absence of anomalies, which will be discussed in Section 2.4.) However, one cannot add vector meson mass terms to the Lagrangian because such terms would break the gauge invariance and lead to a non-renormalizable theory. It therefore appears that the vector measons must be massless and that the forces which they mediate must be long ranged.

This is, of course, desirable for quantum electrodynamics (QED), for which the gauge boson is the photon. The strong and weak interactions are not long ranged, however, and naively they do not seem to fit into the gauge theory framework. It took many years to realize how to use gauge theories to describe the weak and strong interactions. The weak interaction gauge bosons are believed to acquire mass from a spontaneous breakdown of the gauge symmetry in the vacuum. This mechanism is described in Sections 2.3.3 and 2.4. The strong gauge group is not believed to be spontaneously broken. Rather, the same confinement mechanism that presumably prevents quarks from existing as free particles is believed to also prevent the strong gauge bosons from propagating freely (as would be needed to generate a long-range force).

2.3.3 Spontaneous Symmetry Breaking

Renormalizability requires that the Lagrangian (and therefore the equations of motion) of a gauge theory must be exactly invariant under gauge transformations. Since vector boson mass terms are not gauge invariant it appears that the gauge bosons must be exactly massless. There is a loophole to this reasoning, however. Namely, it is possible for the symmetries of the equations of motion of a theory to be broken by the stable solutions, which can pick out a specific direction in the symmetry space. This situation is known as spontaneous symmetry breaking [2.12-22]. A ferromagnetic is a simple example of spontaneous symmetry breaking (SSB). The equations of motion are rotationally invariant, but the spins in a real ferromagnetic are aligned in a definite direction. If SSB occurs in a gauge theory the associated gauge bosons will acquire masses. [2.23-24]

Spontaneous symmetry breaking occurs when the lowest energy state of a theory possesses a nonzero distribution of the charge associated with a symmetry generator. A gauge boson propagating through this vacuum state will constantly interact with this charge and will develop an effective mass proportional to the vacuum expectation value of the charge. The associated force will be shielded and will therefore become short ranged in much the same way that the Coulomb force becomes short ranged in a plasma due to shielding effects [2.24].

The Higgs mechanism [2.25-28] is a simple explicit model for implementing spontaneous symmetry breaking. One introduces a set of spin-0 fields into the theory which transform in a nontrivial way under the gauge symmetry. If the vacuum expectation value (VEV) of one of these fields is nonzero (this is essentially a Bose condensation), then all of the symmetry generators for which this field has a nonzero charge will be spontaneously broken and the associated gauge bosons will be massive.

The Higgs Mechanism

Consider a gauge theory with n Hermitian scalar fields ϕ_a, $a=1,\ldots,n$, arranged in a column vector ϕ, as well as fermion fields represented by ψ. The most general renormalizable Lagrangian invariant under a local symmetry group G is then [2.8]

$$\mathscr{L} = -\frac{1}{4} F^{i}_{\mu\nu} F^{i\mu\nu} + \bar{\psi}(i \not{D} - m_o)\psi + \frac{1}{2}(D^{\mu}\phi)(D_{\mu}\phi) + \bar{\psi}\Gamma^{a}\psi\phi_{a} - V(\phi) , \tag{2.70}$$

where the fermion mass $m_o = m_{oL} P_L + m_{oR} P_R$ and the Yukawa couplings $\Gamma^a = \Gamma^a_L P_L + \Gamma^a_R P_R$ are matrices in the space of fermion group indices and where the potential $V(\phi)$ is a fourth order polynomial in the scalar fields. (Terms higher than fourth order in ϕ or derivatives, other than in the kinetic energy terms, would spoil the renormalizability of the theory.) m_o, Γ^a, and $V(\phi)$ must be chosen to be invariant under global transformations G. The VEV of ϕ is

$$v \equiv \langle 0|\phi|0\rangle = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \tag{2.71}$$

where $|0\rangle$ is the vacuum state and $v_a \equiv \langle 0|\phi_a|0\rangle$. Some or all of the v_a can be zero. v is determined by the condition that the effective potential [2.29-30,2.1] for ϕ, which at the tree diagram level in an expansion in the number of loops is just $V(\phi)$, be minimized at $\phi = v$. That is (at tree level)

$$\left.\frac{\partial V}{\partial \phi_a}\right|_{\phi=v} = 0. \tag{2.72}$$

One can define n quantum fields

$$\hat{\phi}_a \equiv \phi_a - v_a \tag{2.73}$$

with zero VEV. Re-expressing $V(\phi)$ in terms of $\hat{\phi}$ one has the scalar mass terms

$$\mathscr{L}_{mass} = -\frac{\mu^2_{ab}}{2}\hat{\phi}_a\hat{\phi}_b , \tag{2.74}$$

where

$$\mu^2_{ab} = \left.\frac{\partial^2 V}{\partial\phi_a\,\partial\phi_b}\right|_{\phi=v} . \tag{2.75}$$

Under an infinitesimal global transformation,

$$V(\phi) \to V(\phi) + \delta V(\phi) , \tag{2.76}$$

where

$$\delta V(\phi) = \frac{\partial V}{\partial\phi_a}\delta\phi_a = -i\frac{\partial V}{\partial\phi_a}\vec{\beta}\cdot\vec{L}_{ab}\,\phi_b . \tag{2.77}$$

Hence, invariance of V requires $\delta V = 0$, or

$$\frac{\partial V}{\partial\phi_a}L^i_{ab}\,\phi_b = 0 . \tag{2.78}$$

Differentiating (2.78) with respect to ϕ_c and evaluating at $\phi = v$ then yields

$$\mu^2_{ab}(L^i v)_b = 0 , \quad i = 1,\ldots,N . \tag{2.79}$$

Let us label the generators so that

$$\begin{aligned} L^i v &= 0 , \quad i = 1,\ldots,M \\ L^i v &\neq 0 , \quad i = M+1,\ldots,N \end{aligned} \tag{2.80}$$

The subgroup G' of G generated by $T^1 \ldots T^M$ therefore leaves the vacuum invariant ($T^i|0\rangle = 0$). The generators $T^{M+1} \ldots T^N$ do not leave the vacuum invariant ($T^i|0\rangle \neq 0$), so that G is spontaneously broken down to G'. Of course, one can have the special cases M = 0 (G completely broken) and M = N (G' = G). If the original symmetry group G had been merely a global symmetry, then the Nambu-Goldstone theorem [2.19-22] states that for each of the N-M spontaneously broken generators there will exist a massless spin-0 particle in the physical spectrum of the theory. This can be seen (in tree approximation) from (2.79): the scalar mass matrix μ^2_{ab} has N-M eigenvectors $L^i v$, $i = M + 1,\ldots,N$ (which can be shown to be linearly independent [2.1]) with eigenvalue zero. μ^2_{ab} also has $p = n - (N-M)$ (generally) non-zero eigenvalues, corresponding to p (generally) massive scalar particles.

Nambu-Goldstone (N-G) bosons do not appear to exist in nature. Fortunately, when G is a gauge symmetry, the two problems of the unwanted Nambu-Goldstone bosons and the unwanted massless gauge bosons can cure each other [2.25-28]; instead of existing as a massless spin-0 particle, the degree of freedom carried by the N-G boson manifests itself as the longitudinal spin component of the gauge boson, which has in the process acquired a mass. (One says that the N-G boson has been "eaten"). To see this, it is convenient to write [2.28]

$$\phi = e^{i \sum_{i=M+1}^{N} \xi^i L^i} \begin{pmatrix} v_1 + \eta_1 \\ v_2 + \eta_2 \\ \vdots \\ v_p + \eta_p \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \equiv e^{i \sum \xi^i L^i} (v+\eta) , \qquad (2\text{-}81)$$

where the n quantum fields $\hat{\phi}_a \equiv \phi_a - v_a$ have been re-expressed in terms of N-M Nambu-Goldstone fields ξ^i, $i = M+1,\ldots,N$ and $p = n - (N-M)$ physical spin-0 fields known as Higgs particles. Some of the v_a in (2.81) may be zero. To display the physical particles of the theory one can then make the specific gauge transformation defined by

$$U(\xi) = e^{-i \sum\limits_{i=M+1}^{N} \xi^i L^i} , \tag{2.82}$$

so that

$$\phi \to \phi' = U\phi = v + \eta = \begin{pmatrix} v_1 + \eta_1 \\ \vdots \\ v_p + \eta_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.83}$$

represents p physical spin-0 particles and

$$A^i_\mu \to A^{i'}_\mu = A^i_\mu - \frac{1}{g}\partial_\mu \xi^i + O(\xi^2) , \tag{2.84}$$

$$i = M+1,\ldots N$$

represent N-M massive gauge bosons ($A^1_\mu,\ldots,A^M_\mu$ remain massless). In this gauge the boson masses can be read off from the ϕ kinetic energy terms. One has (dropping the primes)

$$\begin{aligned} \tfrac{1}{2}(D^\mu\phi)(D_\mu\phi) &= (v+\eta)^T\,(\overleftarrow{\partial}^\mu + ig\vec{A}^\mu\cdot\vec{L})(\vec{\partial}_\mu - ig\vec{A}_\mu\cdot\vec{L})(v+\eta) \\ &= \tfrac{1}{2}v^T L^i L^j v\, A^{i\mu} A^j_\mu + \text{interaction terms} \\ &= \tfrac{1}{2}M^2_{ij}\, A^{i\mu} A^j_\mu , \end{aligned} \tag{2.85}$$

where the gauge boson mass matrix is

$$
\begin{aligned}
M^2_{ij} = M^2_{ji} &= g^2 v^T L^i L^j v \\
&= g^2 \langle v|L^i L^j v\rangle \qquad (2.86) \\
&= g^2 \langle L^i v|L^j v\rangle
\end{aligned}
$$

and where $\langle x|y\rangle \equiv \Sigma\, x^*_a y_a$. M^2_{ij} has M zero eigenvalues, corresponding to the M unbroken generators, and N-M non-zero eigenvalues.

The Feynman Rules and the R_ξ Gauges

The Feynman rules for a spontaneously broken gauge theory can be read off from (2.70), with ϕ replaced by $v + \hat{\phi}$ (note that in a general gauge $\hat{\phi}$ contains the N-G fields). In particular

$$
\begin{aligned}
\tfrac{1}{2} D^\mu \phi D_\mu \phi = {} & \tfrac{1}{2} D^\mu \hat{\phi} D_\mu \hat{\phi} \\
& + \frac{M^2_{ij}}{2} A^{i\mu} A^j_\mu \\
& + ig \langle v|L^i \partial^\mu \hat{\phi}\rangle A^i_\mu \qquad (2.87) \\
& + g^2 \langle v|L^i L^j \hat{\phi}\rangle A^{i\mu} A^j_\mu \, .
\end{aligned}
$$

The first term leads to the same interaction vertices as in the non-spontaneously broken case (Fig. 2.2); the second term is the vector mass term; the third term will be cancelled by terms added to the Lagrangian to fix the gauge [2.31,2.1]; the last term in (2.87) represents a new three-point vertex illustrated in Fig. 2.3.

Similarly,

$$\bar{\psi}\Gamma^a \psi \phi_a = \bar{\psi}\Gamma^a \psi v_a + \bar{\psi}\Gamma^a \psi \hat{\phi}_a \, . \qquad (2.88)$$

The first term contributes to the fermion masses, so that the total fermion mass matrix is

$$m = m_o - \Gamma^a v_a \ . \tag{2.89}$$

The second term yields the Yukawa couplings of the $\hat{\phi}$. The Higgs potential $V(v+\hat{\phi})$ generates three and four point vertices for the $\hat{\phi}$, as well as the $\hat{\phi}$ mass terms.

The form of the propagators for the boson fields depends on the gauge, although physical quantities such as S-matrix elements are gauge-independent[2.15-17]. It is convenient to work in a special class of gauges called the R_ξ gauges[2.31,2.1,2.16-17]; ξ, which runs from 0 to ∞, parametrizes the gauge. The quantization procedure for the R_ξ gauges introduces additional terms into the effective interaction which modify the structure of the vector and scalar propagators and cancel the mixed $\partial^\mu\hat{\phi}A_\mu$ term in (2.87). There are other terms which can be represented by a set of N "ghost" particles ω_i, $i = 1,\ldots,N$, which are complex scalar fields satisfying Fermi statistics [2.32]. They are needed to ensure unitarity and renormalizability. The ghost fields do not represent physical particles: they occur only as internal lines in Feynman diagram loops. The ghost vertices, given by an effective Lagrangian [2.31,2.1,2.17]

$$\begin{aligned} \mathscr{L}_{\text{ghost}} &= -g(\partial^\mu\omega_i^\dagger)\ c_{ijk}\ \omega_j\ A_\mu^k \\ &\quad - \frac{g^2}{\xi}\,\omega_i^\dagger\omega_j\ \langle v|L^i L^j\hat{\phi}\rangle \ , \end{aligned} \tag{2.90}$$

are shown in Fig. 2.4. There is a factor of -1 for each closed ghost loop.

The propagators for vector and ghost particles in the R_ξ gauge are [2.31,2.1,2.17]

$$i\, D^V_{\mu\nu}(k) = -i\left[g_{\mu\nu} - \frac{k_\mu k_\nu (1-\frac{1}{\xi})}{k^2 - \frac{M^2}{\xi}}\right] \frac{1}{k^2-M^2} \tag{2.91}$$

and

$$i\, D^G(k) = i\, \frac{1}{k^2 - \frac{M^2}{\xi}} \quad , \tag{2.92}$$

respectively. Both are N×N matrices. The scalar propagator

$$i\, D^\phi(k) = i\, P\, \frac{1}{k^2 - \frac{M^2}{\xi}} + (1-P)\, \frac{i}{k^2-\mu^2} \tag{2.93}$$

is an n×n matrix. P, the projection operator onto the N-M dimensional space of Nambu-Goldstone fields (spanned by the vectors $L^i v$, i = M + 1,...,N), is given explicitly by

$$P_{ab} = g^2 (L^i v)_a \left[\frac{1}{M^2}\right]_{ij} (L^j v)^T_b \ , \tag{2.94}$$

and the first term in (2.93) is defined as

$$\left(P\, \frac{1}{k^2 - \frac{M^2}{\xi}}\right)_{ab} \equiv g^2 (L^i v)_a \left(\frac{1}{M^2}\, \frac{1}{k^2 - \frac{M^2}{\xi}}\right)_{ij} (L^j v)^T_b \ . \tag{2.95}$$

D^ϕ can be rewritten

$$D^\phi(k)_{ab} = \left[\frac{i}{k^2-\mu^2}\right]_{ab} + \frac{ig^2}{\xi k^2} (L^i v)_a \left(\frac{1}{k^2 - \frac{M^2}{\xi}}\right)_{ij} (L^j v)^T_b \ , \tag{2.96}$$

where $P\mu^2 = 0$ has been used.

Note that the poles in $D^V_{\mu\nu}$, D^G, and D^ϕ at the gauge-dependent point

$k^2 = M^2/\xi$ cancel in S-matrix elements [2.31,2.1,2.17].

The special gauge $\xi = 0$ is known as the U or unitary gauge. The unitary gauge is very convenient when working in the tree approximation because the ghost fields and N-G fields drop out; hence, one needs to only consider the exchanges of physical vector bosons and physical Higgs particles. The U gauge is not so convenient when one considers higher order diagrams. This is because the vector propagator

$$D^V_{\mu\nu}(k) = -i[g_{\mu\nu} - \frac{k_\mu k_\nu}{M^2}] \frac{1}{k^2-M^2} \qquad (2.97)$$

induces severe ultraviolet divergences which must be handled very carefully [2.8]. Weinberg has also shown [2.8] that there is an effective multi-scalar interaction in the U gauge

$$\mathscr{L}_{eff} = -i\ \delta^4(0)\ \mathrm{Tr}\ \ell n(I+J)\ , \qquad (2.98)$$

where

$$J_{ij} = g^2\left(\frac{1}{M^2}\right)_{ik} \langle v|L^k L^j|\hat{\phi}\rangle\ . \qquad (2.99)$$

The trace and matrices in (2.98) and (2.99) are restricted to the N-M dimensional subspace of broken generators of G. $\mathscr{L}_{eff}$ is a remnant of the ghost loops which survives as $\xi \to 0$ because of the factors ξ^{-1} in the ghost-ghost-scalar vertices which cancel the zeroes in the ghost propagators. $\mathscr{L}_{eff}$ is necessary to cancel divergences in gauge boson loops.

R_ξ gauges for $\xi \neq 0$ are referred to as renormalizable gauges because the vector propagator is better behaved at large momentum. In particular, the gauge $\xi = 1$, which is convenient for calculations, is known as the t'Hooft-Feynman gauge [2.15], while the gauge $\xi = \infty$ is the Landau or R gauge. It is usually wise to calculate for arbitrary ξ in order to verify that the ξ dependence drops out of observable quantities.

A general formalism for treating spontaneously broken gauge theories has been given by Weinberg [2.8]. Renormalizability is shown in [2.15-17].

The Higgs mechanism for spontaneous symmetry breaking requires the introduction of elementary spin-0 fields into the theory. It has often been speculated that spontaneous symmetry breaking may come about without the introduction of elementary scalars. In this case a bound state N-G boson would presumably take the place of the elementary fields. This possibility is sometimes referred to as dynamical symmetry breaking. Some recent speculations along these lines will be briefly described in Section 2.5.3.

2.4 The Weak and Electromagnetic Interactions

2.4.1 Quantum Electrodynamics [2.37]

The electromagnetic interactions are successfully described by quantum electrodynamics (QED), which is a U_1 gauge theory of the type described in (2.40-2.41). The coupling g is replaced by $e > 0$, the charge of the positron, and q_a is the electric charge of particle a in units of e. Under a gauge transformation

$$\begin{aligned} e^- &\to e^{+i\,\beta(x)}\, e^- \\ \nu_e &\to \nu_e \\ u &\to e^{-2/3i\,\beta(x)}\, u \\ d &\to e^{+1/3i\,\beta(x)}\, d\ , \end{aligned} \tag{2.100}$$

where the transformations of u and d are independent of color. The heavier particles transform in a similar way. The amplitude to emit or absorb a photon (γ) is proportional to i e J^{μ}_{EM}, where the electromagnetic current is

$$J^{\mu}_{EM} = -\bar{e}\,\gamma^{\mu}e + \frac{2}{3}\,\bar{u}\gamma^{\mu}u - \frac{1}{3}\,\bar{d}\gamma^{\mu}d + \ldots \qquad (2.101)$$

Some typical electromagnetic diagrams are shown in Fig. 2.5. Of course, the photon is electrically neutral and the vertices are diagonal in the fermion type. Hence, not only electric charge, but also individual particle numbers are conserved by QED. That is, electron number, ν_e number, I^3, hypercharge, strangeness, charm, etc. are all conserved because of global symmetries in QED. Also, J^{μ}_{EM} is non-chiral so that parity is conserved. QED is even under charge conjugation and time reversal transformations.

The U_1 symmetry is not spontaneously broken, so the photon is massless and the Coulomb force is long-ranged. (The experimental upper limit on the photon mass is 6×10^{-22} MeV [2.34]). The electromagnetic fine structure constant is

$$\alpha_e = \frac{e^2}{4\pi} = \frac{1}{137.036} \quad . \qquad (2.102)$$

This is actually the coupling measured at $q^2 = 0$, where q is the four momentum of the photon. For $q^2 \neq 0$ the effective α_e increases logarithmically with q^2. This is due to momentum dependent vacuum polarization effects which will be discussed in Section 2.5 and Chapters 3 and 4.

QED is reviewed and compared with experiment in refs. [2.35-37].

2.4.2 The Weak Interactions [2.38]

The weak interactions, which are responsible for β decay, muon decay, most hyperon decays, etc., are characterized as being very weak and very short ranged. The limit on the range is approximately $R < 10^{-2}$ F, which corresponds to an intermediate vector boson mass of more than 20 GeV. Much of the original knowledge of the weak interactions was obtained because they do not conserve such quantum numbers as strangeness, charm, and I^3 (the third component of isospin), which are respected by the strong and electromagnetic interactions. The weak interactions do not conserve parity (P), which means they distinguish between left and right helicity particles, and they are not charge-conjugation (C) invariant. P and C are violated maximally by the charged current part of the weak interactions, but the product CP is approximately conserved. A very small violation of CP, with strength $\simeq 10^{-3}$ of the weak interaction strength, has been observed in kaon decays, but it is not certain whether this is due to a small piece of the weak interactions or to a new very weak interaction.

Prior to the discovery of neutral currents, most known aspects of the weak interactions (with the exception of CP violation and,possibly, some aspects of non-leptonic hyperon decay) could be described by a modern version of the old Fermi theory of weak interactions, generalized to include such effects as parity violation and strangeness changing decays.

The Fermi theory described the weak interactions in terms of a four-fermion(zero-range) effective Lagrangian,

$$\mathscr{L}_W = \frac{1}{2} \frac{G_F}{\sqrt{2}} \left(J^\mu(x)\, J^\dagger_\mu(x) + J^{\mu\dagger}(x)\, J_\mu(x) \right) , \qquad (2.103)$$

where $G_F \simeq 1.027 \times 10^{-5} M_p^{-2}$ [2.39] is the Fermi constant and M_p is the proton mass. It took many years to determine the form of the weak current $J^\mu(x)$. The final form due to Cabibbo [2.40], Feynman and Gell-Mann [2.41], Sudarshan and Marshak [2.42], Lee and Yang [2.43], and many others [2.38], written in terms of quark and lepton fields, is

$$\begin{aligned} J_\mu &= \bar{e}\, \gamma_\mu\, (1 + \gamma_5)\, \nu_e + \bar{\mu}\, \gamma_\mu (1 + \gamma_5)\, \nu_\mu \\ &\quad + \cos\theta_c\, \bar{d}\, \gamma_\mu (1 + \gamma_5)\, u \\ &\quad + \sin\theta_c\, \bar{s}\, \gamma_\mu (1 + \gamma_5)\, u \\ &= 2(\bar{e}_L\, \gamma_\mu\, \nu_{eL} + \bar{\mu}_L\, \gamma_\mu\, \nu_{\mu L} \\ &\quad + \cos\theta_c\, \bar{d}_L\, \gamma_\mu\, u_L + \sin\theta_c\, \bar{s}_L\, \gamma_\mu\, u_L), \end{aligned} \tag{2.104}$$

where the Cabibbo angle θ_c measures the relative strength of the strangeness-changing and strangeness conserving interactions. From hyperon and kaon decays one has [2.39] $\sin\theta_c \simeq 0.228$. Note that the currents in (2.103) change the electric charge of the particles by $\pm e$, so (2.103) describes "charged current" interactions. J_μ contains equal admixtures of vector and axial vector currents. This implies that C and P are maximally violated by $\mathscr{L}_W$. CP, however, is conserved.

P, C, and CP

For later reference, it will be useful to review the formalism of P, C, and CP at this point. Let $\psi(x) = \psi(\underset{\sim}{x}, t)$ represent a four-component Dirac field, which is an operator which annihilates a particle or creates an antiparticle. One can write ψ as the sum of two two-components Weyl spinors

$$\psi = \psi_L + \psi_R = \frac{1+\gamma_5}{2}\psi + \frac{1-\gamma_5}{2}\psi = P_L\psi + P_R\psi . \qquad (2.105)$$

To the extent that the particle's mass can be ignored $\psi_L(\psi_R)$ annihilates a left (right) handed particle or creates a right (left) handed antiparticle. Amplitudes to create or annihilate a particle with the wrong helicity are proportional to the particle's mass. Under parity or space reflection (P), ψ transforms as [2.33]

$$\psi(\underset{\sim}{x}, t) \underset{P}{\rightarrow} \gamma_o\, \psi(-\underset{\sim}{x}, t), \qquad (2.106)$$

so that

$$\begin{aligned} &\psi_{L,R}(\underset{\sim}{x}, t) \underset{P}{\rightarrow} \gamma_o\, \psi_{R,L}(-\underset{\sim}{x}, t) \\ &\bar{\psi}_{L,R} \equiv (\psi_{L,R})^\dagger \gamma_o \underset{P}{\rightarrow} \bar{\psi}_{R,L}(-\underset{\sim}{x}, t)\, \gamma_o . \end{aligned} \qquad (2.107)$$

(These transformations and those given for charge conjugation can be generalized by including a phase factor on the right hand side.) The Cabibbo current (2.104) involves only left-handed fields (i.e., only left-handed particles and right handed antiparticles interact via $\mathscr{L}_W$). But under P,

$$\bar{\psi}_{1L}(\underset{\sim}{x}, t)\gamma_\mu\, \psi_{2L}(\underset{\sim}{x}, t) \underset{P}{\rightarrow} \bar{\psi}_{1R}(-\underset{\sim}{x}, t)\gamma^\mu\, \psi_{2R}(-\underset{\sim}{x}, t) . \qquad (2.108)$$

That is, a left handed current is transformed into a right handed current. Hence, $\mathscr{L}_W$ is not invariant under P. Under charge conjugation (C),

$$\begin{aligned} \psi &\underset{C}{\rightarrow} \psi^c \equiv C\, \bar{\psi}^T \\ \bar{\psi} &\underset{C}{\rightarrow} \bar{\psi}^c \equiv -\psi^T C^{-1} , \end{aligned} \qquad (2.109)$$

where ψ^c is the charge conjugate field which annihilates an antiparticle or creates a particle, and C is a Dirac matrix defined by

$$C \gamma_\mu C^{-1} = -\gamma_\mu^T \quad . \tag{2.110}$$

In the representation that I am using for the Dirac matrices [2.33], $C = -C^{-1} = -C^\dagger = -C^T = i\gamma^2 \gamma^o$. It is easy to show that

$$\begin{aligned} \psi_{L,R} &\underset{C}{\rightarrow} \psi^c_{L,R} \equiv P_{L,R}\, \psi^c = C\, \bar{\psi}^T_{R,L} \\ \bar{\psi}_{L,R} &\underset{C}{\rightarrow} \overline{\psi^c_{L,R}} = -\, \psi^T_{R,L}\, C^{-1} \end{aligned} \tag{2.111}$$

It follows that

$$\begin{aligned} \psi_{L,R} &= C\, \overline{\psi^c_{R,L}}^T \\ \bar{\psi}_{L,R} &= -\psi^{cT}_{R,L} C^{-1} \end{aligned} \tag{2.112}$$

It should be emphasized that ψ^c_L (ψ^c_R) is the field which annihilates a left (right) handed antiparticle or creates a right (left) handed particle. Under charge conjugation,

$$\bar{\psi}_{1L}\, \gamma_\mu\, \psi_{2L} \underset{C}{\rightarrow} -\bar{\psi}_{2R}\, \gamma_\mu\, \psi_{1R} \tag{2.113}$$

(where the anticommutativity of the fields has been used), so $\mathscr{L}_W$ is not invariant under charge conjugation.

Under the product CP,

$$\psi_{L,R} \underset{CP}{\rightarrow} \gamma_o\, \psi^c_{R,L} \tag{2.114}$$

so that

$$\bar{\psi}_{1L}(\underset{\sim}{x},t)\ \gamma_\mu\ \psi_{2L}(\underset{\sim}{x},t) \underset{CP}{\rightarrow} -\bar{\psi}_{2L}(-\underset{\sim}{x},t)\ \gamma^\mu\ \psi_{1L}(-\underset{\sim}{x},t) \quad . \tag{2.115}$$

This implies that the weak current transforms as

$$J_\mu(\underset{\sim}{x},\ t) \underset{CP}{\rightarrow} - J^\mu(-\underset{\sim}{x},\ t)^\dagger \ , \tag{2.116}$$

under CP transformations, so that

$$\mathscr{L}_W(\underset{\sim}{x},\ t) \underset{CP}{\rightarrow} \mathscr{L}_W(-\underset{\sim}{x},\ t); \tag{2.117}$$

the action $\int d^4x \ \ \mathscr{L}_W(\underset{\sim}{x},\ t)$ is invariant.

It is fairly common to regard ψ_L and ψ_R as fundamental fields. Then ψ_L^c and ψ_R^c are not independent: they are related to $\psi_{L,R}^\dagger$ by (2.111). One could just as well take any of the pairs $(\psi_L,\ \psi_L^c)$, $(\psi_R,\ \psi_R^c)$, or $(\psi_R^c,\ \psi_L^c)$ as fundamental , however. For example, it is conventional to express the charged current J_μ in (2.104) in terms of the left handed operators ψ_L, but it could just as well be written in terms of the charge conjugate fields ψ_R^c using (2.112). Thus,

$$\bar{e}_L \ \gamma_\mu \ \nu_{eL} \ = - \ \overline{\nu^c_{eR}} \ \gamma_\mu \ e^c_R \ . \tag{2.118}$$

In discussing grand unified theories it will generally prove convenient to express the various currents in terms of the left-handed fields ψ_L and ψ_L^c.

The Fermi Theory

Let us return to a discussion of the Fermi theory. The Lagrangian $\mathscr{L}_W$ was phenomenologically successful, in that it correctly described all known charged current weak processes (before the discovery of the charmed quark) except CP violation and possibly the non-leptonic kaon and hyperon decays. Some typical diagrams for processes described by $\mathscr{L}_W$ are shown in Fig. 2.6.

However, the Fermi theory could not be considered to be exact because it violated unitarity at high energies [2.44]. To see this, consider the cross section calculated from $\mathscr{L}_W$ for $\nu_e e^- \to \nu_e e^-$. It is [2.38]

$$\sigma_{tot}(s) = \frac{G_F^2\, s}{\pi} \tag{2.119}$$

where $s = 4\omega^2$ is the square of the total center of mass energy. But $\mathscr{L}_W$ describes a point interaction, so it can only produce S wave scattering. Furthermore, only one helicity state is involved (neglecting m_e). Hence, σ_{tot} should satisfy the unitarity limit

$$\sigma_{tot}(s) < \frac{4\pi}{s} \tag{2.120}$$

Clearly, this unitarity bound is violated for

$$\omega = \frac{\sqrt{s}}{2} > \frac{1}{2}\left(\frac{2\pi}{G_F}\right)^{1/2} \simeq 300 \text{ GeV} ; \tag{2.121}$$

the Fermi theory must break down before this energy. One could attempt to unitarize the theory by computing higher order diagrams in $\mathscr{L}_W$. However, this runs into the difficulty that $\mathscr{L}_W$ describes a non-renormalizable field theory: there are severe and unacceptable ultraviolet divergences in the Feynman integrals (which are closely related to the increase of σ_{tot} at high energy).

The Intermediate Vector Boson Theory

Another possibility, known as the Intermediate Vector Boson (IVB) theory [2.38], is that the four fermion interaction is really just a low energy approximation to a finite range interaction. This interaction is mediated by electrically-charged massive vector particles $W^\pm$ known as intermediate vector bosons, which couple to the Cabibbo current by the interaction

$$\mathscr{L} = \frac{g}{2\sqrt{2}} (J^{\mu} W_{\mu}^{-} + J^{\mu\dagger} W_{\mu}^{+}) \ . \tag{2.122}$$

Typical diagrams are shown in Figure 2-6. The IVB propagator is

$$D_{\mu\nu}^{V}(k) = \frac{-g_{\mu\nu} + \frac{k_{\mu}k_{\nu}}{M_W^2}}{k^2 - M_W^2} \xrightarrow[|k^2| \ll M_W^2]{} \frac{g_{\mu\nu}}{M_W^2} \tag{2.123}$$

Hence, for momentum transfers small compared to M_W the IVB theory reproduces the results of the Fermi theory for

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} \tag{2.124}$$

The IVB theory resembles QED in that the interactions in both are mediated by the exchange of vector particles. Unlike the photon, however, the IVB is massive, electrically charged, and couples only to left-handed particles and right-handed antiparticles.

Unfortunately, the IVB theory fails to produce a unitary renormalizable theory. The high energy behavior of $\nu_e e^- \to \nu_e e^-$ is no longer such a severe problem (the amplitude is no longer purely S-wave) but other amplitudes, such as that for $e^+e^- \to W^+W^-$, with the $W^\pm$ longitudinally polarized, again violate unitary for $\sqrt{s} \gtrsim G_F^{-1/2}$ [2.45]. This is closely related to the fact that higher order Feynman diagrams are still badly divergent because of the $k_\mu k_\nu$ term in the W propagator, leading to a nonrenormalizable theory. The problem is that unlike QED, the IVB theory is not a gauge theory. It is simply a vector meson theory with an elementary mass term in the Lagrangian.

Renormalizable Theories

One would therefore like to embed the Fermi theory into a renormalizable field theory. One interesting possibility, originally due to Kummer and Segré [2.46] is that the weak interactions are mediated by spin-0 bosons, but that the Yukawa couplings are such that the transitions involving light leptons and hadrons require the exchange of two bosons. At low energy the two-boson exchange amplitude can mimic a vector-axial vector interaction.

Another approach, which is used in the standard model, is to modify the IVB theory by incorporating it into a spontaneously broken gauge theory. The interaction (2.122) between $W^{\pm}$ and the weak current will be obtained in a gauge theory if the charges

$$Q^{-} \equiv \int d^3x \, J_0(\underset{\sim}{x}, t) \tag{2.125}$$

and $Q^{+} \equiv (Q^{-})^{\dagger}$ are generators of the gauge group G. However, if Q^{-} and Q^{+} are generators of G, then so is

$$[Q^{-}, Q^{+}] = 2 \int d^3x [e^{\dagger}(1 + \gamma_5)e - \nu^{\dagger}(1 + \gamma_5)\nu] \; , \\ + \ldots \tag{2.126}$$

which is the charge associated with an electrically neutral current. Hence, embedding the IVB theory into a gauge theory will require the existence of at least one neutral current and an associated neutral gauge boson. (It is possible for this to be the photon, but in most models this is not the case.) The interactions mediated by the new boson will cancel many of the divergences in the IVB theory. Furthermore, the troublesome $k_{\mu}k_{\nu}$ term in the vector propagators will now be effectively unobservable due to gauge invariance, which means that one can work in the renormalizable gauges described in Section 2.3.

Of course, the $W^{\pm}$ and neutral boson masses must now be generated by spontaneous symmetry breaking.

It is possible to write down gauge theories for the weak interactions alone [2.47]. However, in the GWS model, to which I now turn, the weak and electromagnetic interactions are both pieces of a larger unifying gauge group, which includes not only QED and the charged current weak interactions, but also a new neutral current weak interaction.

2.4.3 The Glashow-Weinberg-Salam Model [2.44]

History

The Glashow-Weinberg-Salam (GWS) model of weak interactions is actually due to many people. In 1957 Schwinger [2.48] proposed a model with a charge triplet $W^{\pm,0}$ of vector bosons. He identified the W^0 with the photon, and the $W^{\pm}$ possessed tensor rather than axial vector couplings. In 1958 Bludman [2.49] proposed an SU_2 gauge theory of weak interactions. Bludman identified the neutral gauge boson W^0 predicted by his model with a new neutral current interaction. In 1961 Glashow [2.50] unified the weak and electromagnetic interactions in a gauge theory based on the group $SU_2 \times U_1$. A similar model was later considered by Salam and Ward [2.51]. Subsequently, Weinberg [2.52] and Salam [2.53] improved the model by suggesting that the vector bosons could acquire mass via the Higgs mechanism [2.25-28]. Weinberg suggested that the theory might be renormalizable; the proof of renormalizability was given several years later by 't Hooft [2.15], 't Hooft and Veltman [2.16], and Lee and Zinn-Justin [2.17]. The original Weinberg-Salam model described only the weak and electromagnetic interactions of leptons. A naive extension of the model to include quarks predicted the existence of substantial strangeness-changing neutral currents, in disagreement with experiment. However, Weinberg showed [2.54] that hadrons could be incorporated in the model by implementing a mechanism due to Glashow, Iliopoulos, and Maiani (GIM) [2.55]. The GIM mechansim involved the introduction of a fourth quark, called the charmed quark. Strangeness-changing neutral currents associated with the commutator $[Q^-, Q^+]$ in (2.126) are cancelled by similar currents associated with the charged currents involving the charmed quark. The Weinberg-Salam model supplemented with the GIM mechanism therefore predicted both the existence of strangeness conserving neutral currents and of the charmed quark, both of which were subsequently discovered.

Basic Structure

I will now describe the simplest form of the GWS model. Modifications will be considered in the next section.

The GWS model is based on the gauge group $SU_2 \times U_1$. SU_2 is the group of 2×2 unitary matrices with determinant one. It has three generators T^i, $i = 1,2,3$, and therefore three gauge bosons A^i_μ. The structures constants are $c_{ijk} = \varepsilon_{ijk}$, where the Levi-Civita symbol ε_{ijk} is totally antisymmetric in all three indices and $\varepsilon_{123} = +1$. The fundamental 2×2 representation is $L^i = \tau^i/2$, where τ^i is the i^{th} Pauli matrix. The generator and gauge boson of the U_1 subgroup are written Y and B_μ, respectively. g and g' are the gauge coupling constants of the SU_2 and U_1 subgroups.

The GWS model is a chiral model in which parity violation is incorporated by assigning left- and right-handed fermions to different representations. All left-handed fermions transform according to doublet (two dimensional) representations of SU_2, while right-handed fermions are singlets. For this reason, the group is often written SU_{2L} and the generators as T_L^i; although the label L has no group theoretical significance. Both L and R fields transform non-trivially under U_1 transformations. The Y charge assignments of the fermion and boson fields are chosen so that $Q = T_L^3 + Y$ is the electric charge operator. T_L^i and Y are sometimes referred to as the weak isospin and weak hypercharge generators, respectively. (Some authors define $Q = T_L^3 + \frac{1}{2} Y$ and define the U_1 gauge coupling to be $\frac{1}{2} g'$. The two factors of $\frac{1}{2}$ compensate so that the gauge interaction is unchanged.)

The minimal GWS model involves one complex doublet of scalar particles

$$\varphi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} \tag{2.127}$$

where φ^+ and φ^0 have electric charge +1 and 0, respectively. Hence, the Y charge of φ is $y_\varphi = +\frac{1}{2}$. One writes $(T, Y)_\varphi = (2, \frac{1}{2})$, which means that φ transforms according to the two-dimensional representation of SU_2 and $y_\varphi = \frac{1}{2}$. The φ covariant derivative is therefore

$$D_\mu \varphi = (\partial_\mu - ig \frac{\tau^i A_\mu^i}{2} - \frac{ig'}{2} B_\mu)\varphi \ . \tag{2.128}$$

The part of the Lagrangian involving the gauge and Higgs fields is

$$\mathscr{L}_{VM} + \mathscr{L}_\varphi = -\frac{1}{4} F_{\mu\nu}^i F^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + (D^\mu \varphi)^\dagger D_\mu \varphi - V(\varphi) \ , \tag{2.129}$$

where

$$\begin{aligned} F_{\mu\nu}^i &= \partial_\mu A_\nu^i - \partial_\nu A_\mu^i + g \, \varepsilon_{ijk} A_\mu^j A_\nu^k \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{2.130}$$

and the most general $SU_2 \times U_1$ invariant fourth order Higgs potential is

$$V(\varphi) = +\mu^2 \varphi^\dagger \varphi + \lambda(\varphi^\dagger\varphi)^2 \ . \qquad (2.131)$$

λ must be positive in order for $V(\varphi)$ to be bounded from below.

The fermions to be included in the model include ν_e, e^-, u, and d, which are known collectively as a family or generation, and at least two repetitions which form additional families. The second family contains ν_μ, μ^-, c, and s, and the third consists of ν_τ, τ^-, t, and b. The left-handed quarks are grouped into three (or more) SU_2 doublets

$$q^0_{mL} \equiv \begin{pmatrix} u^0_m \\ d^0_m \end{pmatrix}_L \qquad (2.132)$$

where m labels the doublet. (The color index has been suppressed. Each q^0_{mL} actually represents three doublets $q^{\alpha 0}_{mL}$, α = R, G, B.) The superscript 0 means "weak interaction basis." That is, u^0_{mL} and d^0_{mL} are the fields that are grouped together in the m^{th} weak doublet and which are therefore changed into each other by the emission or absorption of a gauge boson--they may be linear combinations of quarks of definite mass. The q^0_{mL} have weak hypercharge $y_{qL} = +\frac{1}{6}$. Similarly, the left-handed leptons are placed in doublets

$$\ell^0_{mL} = \begin{pmatrix} \nu^0_m \\ e^0_m \end{pmatrix}_L \qquad (2.133)$$

with $y_{\ell L} = -\frac{1}{2}$. The right-handed fields e^0_{mR}, u^0_{mR}, and d^0_{mR} are SU_2 singlets with weak hypercharges -1, $+\frac{2}{3}$, and $-\frac{1}{3}$, respectively. Right-handed neutrino fields ν^0_{mR} are usually not introduced into the model.

The gauge covariant kinetic energy terms for the fermions are

$$\mathscr{L}_f = \sum_{m=1}^{F} (\bar{q}^0_{mL}\ i\not{D}\ q^0_{mL} + \bar{\ell}^0_{mL}\ i\not{D}\ \ell^0_{mL}$$
$$+ \bar{u}^0_{mR}\ i\not{D}\ u^0_{mR} + \bar{d}^0_{mR}\ i\not{D}\ d^0_{mR} + \bar{e}^0_{mR}\ i\not{D}\ e^0_{mR}) \ , \tag{2.134}$$

where F is the number of families and where

$$D_\mu\ q^0_{mL} = (\partial_\mu - \frac{ig}{2}\vec{\tau}\cdot\vec{A}_\mu - \frac{ig'}{6}\ B_\mu)q^0_{mL}$$

$$D_\mu\ \ell^0_{mL} = (\partial_\mu - \frac{ig}{2}\vec{\tau}\cdot\vec{A}_\mu + \frac{ig'}{2}\ B_\mu)\ell^0_{mL}$$

$$D_\mu\ u^0_{mR} = (\partial_\mu - \frac{2}{3}\ ig'\ B_\mu)u^0_{mR} \tag{2.135}$$

$$D_\mu\ d^0_{mR} = (\partial_\mu + \frac{1}{3}\ ig'\ B_\mu)d^0_{mR}$$

$$D_\mu\ e^0_{mR} = (\partial_\mu + ig'\ B_\mu)e^0_{mR}$$

Bare fermion mass terms are forbidden by the gauge symmetry.

One can easily construct $SU_2 \times U_1$ invariant Yukawa couplings such as

$$\Gamma^d_{mn}\ \bar{q}^0_{mL}\ \varphi\ d^0_{nR} + \text{H.C.} = \Gamma^d_{mn}(\bar{u}^0_{mL}\ \varphi^+ + \bar{d}^0_{mL}\ \varphi^0)d^0_{nR}$$
$$+ \Gamma^{d*}_{mn}\ \bar{d}^0_{nR}(\varphi^{+\dagger}\ u^0_{mL} + \varphi^{0\dagger}\ d^0_{mL}) \tag{2.136}$$

and

$$\Gamma^e_{mn}\ \bar{\ell}^0_{mL}\ \varphi\ e^0_{nR} + \text{H.C.}\ , \tag{2.137}$$

where H.C. means Hermitian conjugate. Terms of the form $\bar{q}^0_{mL}\ \varphi\ u^0_{nR}$, which one might think are required to generate masses for the charge $\frac{2}{3}$ quarks, are forbidden because they violate U_1 invariance (the total Y or Q is not zero). However, an invariant coupling can be constructed in terms of $\varphi^\dagger$ which has $(T, Y)_{\varphi^\dagger} = (2^*, -\frac{1}{2})$. That is, the 2^* representation matrices according to which $\varphi^\dagger$ transform are (2.11)

$$L^i_{2^*} = -L^{iT}_2 = -\frac{\tau^{iT}}{2} \tag{2.138}$$

However, for SU_2, $L^i_{2^*}$ and L^i_2 are equivalent, because

$$i\tau^2\left(-\frac{\tau^{iT}}{2}\right)(i\tau^2)^\dagger = \frac{\tau^i}{2} \tag{2.139}$$

(Analogous statements do not hold for SU_n, $n > 2$.)

Specifically,

$$\tilde{\varphi} \equiv i\tau^2 \varphi^\dagger = \begin{pmatrix} \varphi^{0\dagger} \\ -\varphi^- \end{pmatrix} \tag{2.140}$$

transforms as $(T, Y)_{\tilde{\varphi}} = (2, -\frac{1}{2})$, as can be readily verified from (2.11). Then, one can write Yukawa couplings

$$\Gamma^u_{mn}\, \bar{q}^0_{mL}\, \tilde{\varphi}\, u^0_{nR} + \text{H.C.} = \Gamma^u_{mn}(\bar{u}^0_{mL}\, \varphi^{0\dagger} - \bar{d}^0_{mL}\, \varphi^-)u^0_{nR} + \text{H.C.} \tag{2.141}$$

The final Yukawa interaction is therefore

$$\mathscr{L}_{\text{Yuk}} = \sum_{mn=1}^{F} \Gamma^u_{mn}\, \bar{q}^0_{mL}\, \tilde{\varphi}\, u^0_{nR} + \sum_{mn=1}^{F} \Gamma^d_{mn}\, \bar{q}^0_{mL}\, \varphi\, d^0_{nR}$$
$$+ \sum_{mn=1}^{F} \Gamma^e_{mn}\, \bar{\ell}^0_{mL}\, \varphi\, e^0_{nR} + \text{H.C.}\ , \tag{2.142}$$

where Γ^u, Γ^d , and Γ^e are arbitrary F×F matrices of Yukawa couplings. The total Lagrangian of the model is

$$\mathscr{L} = \mathscr{L}_{\text{VM}} + \mathscr{L}_\varphi + \mathscr{L}_\text{f} + \mathscr{L}_{\text{Yuk}} + \mathscr{L}_{\text{ghost}} \tag{2.143}$$

Spontaneous Symmetry Breaking

Let us now consider the minimization of the Higgs potential $V(\varphi)$. The complex doublet φ can be decomposed into two Hermitian doublets φ_R and φ_I by $\varphi = \frac{1}{\sqrt{2}}(\varphi_R + i\varphi_I)$, where

$$\varphi_R = \begin{pmatrix} \varphi_R^1 \\ \varphi_R^2 \end{pmatrix} \qquad \varphi_I = \begin{pmatrix} \varphi_I^1 \\ \varphi_I^2 \end{pmatrix} . \tag{2.144}$$

Since $V(\varphi)$ depends only on

$$\varphi^\dagger \varphi = \frac{1}{2}[(\varphi_R^1)^2 + (\varphi_R^2)^2 + (\varphi_I^1)^2 + (\varphi_I^2)^2] , \tag{2.145}$$

the orientation of $<0|\varphi|0>$ is not determined. By convention one takes

$$\begin{aligned} v_R &\equiv <0|\varphi_R|0> = \begin{pmatrix} 0 \\ \nu \end{pmatrix} \\ v_I &= <0|\varphi_I|0> = 0 , \end{aligned} \tag{2.146}$$

so that

$$v = <0|\varphi|0> = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} . \tag{2.147}$$

ν is real because φ_R is Hermitian. Any other orientation of $<0|\varphi|0>$ can be rotated into this conventional form by a global $SU_2 \times U_1$ transformation. In terms of ν the potential and its derivative are

$$\begin{aligned} V(\nu) &= \frac{1}{2}\mu^2 \nu^2 + \frac{\lambda}{4}\nu^4 \\ V'(\nu) &= \nu(\mu^2 + \lambda \nu^2) = 0 . \end{aligned} \tag{2.148}$$

$V(\nu)$ is illustrated for the two cases $\mu^2 > 0$ and $\mu^2 < 0$ in Fig. 2.7. The minimum occurs at

$$\nu = \begin{cases} 0 \ , \ \mu^2 > 0 \\ \left(\frac{-\mu^2}{\lambda}\right)^{\frac{1}{2}} \ , \ \mu^2 < 0 \ . \end{cases} \tag{2.149}$$

Hence, spontaneous symmetry breaking occurs for $\mu^2 < 0$. (Spontaneous symmetry breaking also occurs for sufficiently small λ at the transition point $\mu^2 = 0$, but in this case loop corrections to the effective potential must be considered [2.29].) In this case,

$$\begin{aligned} &\frac{\tau^i}{2} v \neq 0 \\ &y_\varphi \ v = \frac{1}{2} v \neq 0 \\ &\left(\frac{\tau^3}{2} + y_\varphi\right) v = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} v = 0 \ . \end{aligned} \tag{2.150}$$

Hence, the symmetries associated with the generators T^1, T^2 and $T^3 - Y$ are spontaneously broken. However, the subgroup U_1^{EM} generated by the electric charge operator $Q = T^3 + Y$ is unbroken. Hence, $SU_2 \times U_1$ is broken down to the U_1^{EM} of electromagnetism for $\mu^2 < 0$. We therefore expect one massless gauge boson (the photon) and three massive bosons.

The Higgs field can be written

$$\varphi = \frac{1}{\sqrt{2}} e^{i\Sigma\xi^i L^i} \begin{pmatrix} 0 \\ \nu + \eta \end{pmatrix} \ , \tag{2.151}$$

where the sum is over the broken generators and η is a physical Hermitian Higgs field. The notation is slightly different from (2.81) because of the conventional orientation of ν and the non-Hermitian basis. In the unitary gauge,

$$\varphi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + \eta \end{pmatrix} \quad . \tag{2.152}$$

There is one physical Higgs particle. The other three components of φ have been "eaten" to give masses to three of the gauge bosons.

The gauge boson and Higgs mass matrices can be computed using the formalism of Section 2.3.3 (using the Hermitian basis for φ and $\varphi^\dagger$). However, it is simpler to write out $\mathscr{L}_\varphi$ in the unitary gauge using (2.152) and read off the mass terms. For example,

$$\begin{aligned} V(\varphi) &= \frac{\mu^2}{2} (\nu + \eta)^2 + \frac{\lambda}{4} (\nu + \eta)^4 \\ &= \frac{-\mu^4}{4\lambda} - \mu^2 \eta^2 + \lambda \nu \eta^3 + \frac{\lambda}{4} \eta^4 \ . \end{aligned} \tag{2.153}$$

The η^3 and η^4 terms represent interactions, while the constant plays no role in particle physics (it does affect the energy density of the universe--see Section 2.4.4). The quadratic term is $\mu_\eta^2 \eta^2/2$, where μ_η is the Higgs boson mass; hence,

$$\mu_\eta^2 = -2\mu^2 > 0 \tag{2.154}$$

The φ kinetic energy term is

$$\begin{aligned} (D^\mu \varphi)^\dagger D_\mu \varphi &= \frac{1}{2} (0 \ \nu) [\frac{g}{2} \vec{\tau} \cdot \vec{A}_\mu + \frac{g'}{2} B_\mu]^2 \begin{pmatrix} 0 \\ \nu \end{pmatrix} + \eta \text{ terms} \\ &= M_W^2 W^{+\mu} W_\mu^- + \frac{M_Z^2}{2} Z^\mu Z_\mu \end{aligned} \tag{2.155}$$

where

$$W_\mu^\pm = \frac{A_\mu^1 \mp iA_\mu^2}{\sqrt{2}} \tag{2.156}$$

and

$$Z_\mu = \frac{g' B_\mu - g A_\mu^3}{\sqrt{g^2 + g'^2}} = \sin\theta_W B_\mu - \cos\theta_W A_\mu^3 \tag{2.157}$$

are charged and neutral massive gauge fields with masses

$$M_W^2 = \frac{g^2 \nu^2}{4}$$

$$M_Z^2 = (g^2 + g'^2) \frac{\nu^2}{4} = \frac{M_W^2}{\cos^2\theta_W} \quad , \tag{2.158}$$

and where

$$\tan\theta_W = \frac{g'}{g} \tag{2.159}$$

defines the weak (or Weinberg) angle. The fourth gauge boson

$$A_\mu = \frac{g B_\mu + g' A_\mu^3}{\sqrt{g^2 + g'^2}} = \cos\theta_W B_\mu + \sin\theta_W A_\mu^3 \tag{2.160}$$

is massless. It is the photon associated with the unbroken U_1^{EM} subgroup.

Interactions

The gauge couplings in $\mathscr{L}_f$ can be rewritten in terms of the mass eigenstates $W^\pm$, Z, and A. The result is

$$\begin{aligned} \mathscr{L}_f = {} & \text{kinetic energy terms} \\ & + \frac{g}{2\sqrt{2}} (J_W^\mu W_\mu^- + J_W^{\mu\dagger} W_\mu^+) \\ & + \frac{gg'}{\sqrt{g^2 + g'^2}} J_{EM}^\mu A_\mu - \frac{\sqrt{g^2 + g'^2}}{2} J_Z^\mu Z_\mu \ . \end{aligned} \tag{2.161}$$

The charged weak current is

$$J_W^\mu = \sum_m [\bar{e}_m^0 \gamma^\mu (1 + \gamma_5) \nu_m^0 + \bar{d}_m^0 \gamma^\mu (1 + \gamma_5) u_m^0] \ . \tag{2.162}$$

For momenta small compared to M_W, $\mathscr{L}_f$ leads to an effective four-fermi charged current interaction

$$\mathscr{L}_C^{eff} = \frac{G_F}{2\sqrt{2}} (J_W^\mu J_{W\mu}^\dagger + J_W^{\mu\dagger} J_{\mu W}) \ , \tag{2.163}$$

where the Fermi constant is given by

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} = \frac{1}{2\nu^2} \ . \tag{2.164}$$

This relation determines ν.

The electromagnetic current is

$$\begin{aligned} J_{EM}^\mu &= \sum_m q_m \bar{\psi}_m^0 \gamma^\mu \psi_m^0 \\ &= \sum_m (\tfrac{2}{3} \bar{u}_m^0 \gamma^\mu u_m^0 - \tfrac{1}{3} \bar{d}_m^0 \gamma^\mu d_m^0 - \bar{e}_m^0 \gamma^\mu e_m^0) \ , \end{aligned} \tag{2.165}$$

where in the first line the sum extends over all fermion fields ψ_m^0 of charge q_m. One must take

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}} = g \sin\theta_W \tag{2.166}$$

to be the positron charge.

The model predicts the existence of a massive neutral gauge boson Z^μ which couples to the weak neutral current

$$J_Z^\mu = \sum_m \bar{\psi}_m^0 \gamma^\mu [T_{mL}^3(1 + \gamma_5) - 2q_m \sin^2\theta_W]\psi_m^0 \ , \tag{2.167}$$

where again the sum extends over all the fermions. T_{mL}^3 is the value of T^3 for ψ_{mL}^0. The second term is just $-2 \sin^2\theta_W J_{EM}^\mu$. At small momentum transfer, the neutral current interaction can be described by an effective four-fermion interaction

$$\mathscr{L}_Z^{\text{eff}} = \frac{G_F}{\sqrt{2}} J_Z^\mu J_{\mu Z} \tag{2.168}$$

Two constraints on the unknown parameters g, g', and ν are given by $G_F/\sqrt{2} = 1/2\nu^2$ and $e = g \sin\theta_W$. $\sin^2\theta_W$ can be obtained independently by measuring neutral current processes. M_W and M_Z are then predicted to be

$$\begin{aligned} M_W &= \frac{g\nu}{2} = \left(\frac{\pi\alpha}{\sqrt{2}G_F}\right)^{\frac{1}{2}} \frac{1}{\sin\theta_W} = \frac{37.3 \text{ GeV}}{\sin\theta_W} \\ M_Z &= \frac{M_W}{\cos\theta_W} = \frac{74.6 \text{ GeV}}{\sin 2\theta_W} \end{aligned} \tag{2.169}$$

Higher order weak effects modify the coefficients in (2.169) to 38.53 GeV and 77.06 GeV, respectively [2.56]. Renormalization in the model is considered in [2.57].

Fermion Mass Eigenstates

It still remains to rewrite the gauge interactions in terms of fermion fields of definite mass. In the unitary gauge, the Yukawa interaction becomes

$$\begin{aligned} \mathscr{L}_{\text{Yuk}} &= \sum_{mn} \bar{u}_{mL}^0 \Gamma_{mn}^u \frac{\nu + \eta}{\sqrt{2}} u_{nR}^0 + \text{H.C.} \\ &= \bar{u}_L^0 (M^u + h^u \eta) u_R^0 + \text{H.C.} \ , \end{aligned} \tag{2.170}$$

where similar terms for the d^0's and e^0's are understood and where the $u^0_{mL,R}$ have been placed in column vectors $u^0_{L,R} = (u^0_1, u^0_2, \ldots u^0_F)^T_{L,R}$. The mass matrix M^u and Yukawa coupling matrix h^u are defined by

$$\begin{aligned} M^u_{mn} &= \Gamma^u_{mn} \frac{\nu}{\sqrt{2}} \\ h^u_{mn} &= \frac{1}{\nu} M^u_{mn} = \frac{g}{2M_W} M^u_{mn} \ . \end{aligned} \tag{2.171}$$

They are in general complex and non-Hermitian (unless extra symmetries are added to the theory to constrain the Γ^u_{mn}). Similar definitions hold for M^d, h^d, M^e, and h^e. I will therefore drop the superscripts u, d, and e in the following.

The mass matrices may be "diagonalized" by means of a transformation

$$A_L^\dagger M A_R = M_D = \begin{pmatrix} m_1 & & & & 0 \\ & m_2 & & & \\ & & m_3 & & \\ & & & \ddots & \\ 0 & & & & m_F \end{pmatrix} , \tag{2.172}$$

where M_D is real and diagonal and m_n are the mass "eigenvalues" of M (i.e., $(m_1, m_2, m_3, \ldots) = (m_u, m_c, m_t, \ldots)$ for the charge $\frac{2}{3}$ quarks). A_L and A_R are F×F unitary matrices which are in general different except in the special case that M is Hermitian. A_L and A_R can be almost uniquely determined by observing that $MM^\dagger$ and $M^\dagger M$ are Hermitian F×F matrices and that from (2.172)

$$A_L^\dagger MM^\dagger A_L = A_R^\dagger M^\dagger M A_R = M_D^2 = \begin{pmatrix} m_1^2 & & & 0 \\ & m_2^2 & & \\ & & \ddots & \\ 0 & & & m_F^2 \end{pmatrix} . \tag{2.173}$$

Hence, A_L and A_R can be determined by elementary methods from (2.173).

Define the fields $u_{L,R} = (u_1, u_2, \ldots u_F)^T_{L,R}$ by

$$u_L^0 = A_L^u \, u_L$$
$$u_R^0 = A_R^u \, u_R \; . \tag{2.174}$$

Similar definitions hold for $d_{L,R}$ and $e_{L,R}$. The mass and Yukawa terms are diagonalized when rewritten in terms of u_L and u_R:

$$\begin{aligned} &\bar{u}_L^0 (M^u + h^u \, \eta) u_R^0 + \text{H.C.} \\ &= \bar{u}_L (M_D^u + h_D^u \, \eta) u_R + \text{H.C.} \\ &= \bar{u} (M_D^u + h_D^u \, \eta) u \\ &= (\bar{u} \; \bar{c} \; \bar{t}) \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \left(1 + \frac{g}{2M_W} \eta \right) \begin{pmatrix} u \\ c \\ t \end{pmatrix} , \end{aligned} \tag{2.175}$$

where

$$h_D^u \equiv A_L^{u\dagger} \, h^u \, A_R^u = \frac{g}{2M_W} M_D^u \; . \tag{2.176}$$

In the last line I have specialized to the case of $F = 3$, and have identified $(u_1, u_2, u_3)_{L,R}$ with $(u,c,t)_{L,R}$. The $u_{L,R}$ basis is referred to as the mass eigenstate basis. Similar statements apply to the charge $-\frac{1}{3}$ quarks and to the leptons.

The Yukawa matrices h are proportional to the mass matrices (2.171). Hence, they are made real and diagonal by the same transformations that diagonalize M. This implies that the couplings of the η are diagonal in flavor, that they are scalar (no γ_5's), and that the η couples preferentially to the most massive fermions. However, the Yukawa couplings $gm/2M_W$ are predicted to be extremely weak for all the known fermions because of the m/M_W factor. It should be emphasized that all of these statements would in general be modified if two or more complex Higgs doublets were included in the model (see Section 2.4.4).

Equation (2.173) determines A_L (and A_R) only up to F arbitrary phases. That is, if A_L and A_R satisfy (2.173) then so do $A_L' \equiv A_L K_L$ and $A_R' \equiv A_R K_R$, where

$$K_L = \begin{pmatrix} e^{i\varphi_{1L}} & & & 0 \\ & e^{i\varphi_{2L}} & & \\ & & \ddots & \\ 0 & & & e^{i\varphi_{FL}} \end{pmatrix} . \tag{2.177}$$

A similar definition holds for K_R. The relative phases $\varphi_{iL} - \varphi_{iR}$ are determined from (2.172) by the condition that M_D be real and positive, but the absolute phases are arbitrary. Another way of stating this ambiguity is that one can redefine the phases of the mass eigenstate fields by the transformations

$$\begin{aligned} u_L' &= K_L^\dagger u_L \\ u_R' &= K_R^\dagger u_R . \end{aligned} \tag{2.178}$$

The form of the mass terms and Yukawa couplings is left invariant as long as $K_L = K_R$ (the gauge terms may change).

We are now in a position to reexpress the gauge couplings in terms of the fermion fields of definite mass (i.e., the mass eigenstate basis). The form of the electromagnetic and weak neutral currents are unchanged:

$$
\begin{aligned}
J^{\mu}_{EM} &= \sum_m q_m \, \overline{\psi}_m \, \gamma^{\mu} \, \psi_m \\
&= \frac{2}{3} \, \overline{u} \, \gamma^{\mu} \, u - \frac{1}{3} \, \overline{d} \, \gamma^{\mu} \, d - \overline{e} \, \gamma^{\mu} \, e \\
&\quad + \cdots \\
J^{\mu}_{Z} &= \sum_m \overline{\psi}_m \, \gamma^{\mu} [T^3_{mL}(1 + \gamma_5) - 2q_m \sin^2\theta_W]\psi_m \\
&= \frac{1}{2} \, \overline{u} \, \gamma^{\mu}(1 + \gamma_5)u - \frac{1}{2} \, \overline{d} \, \gamma^{\mu}(1 + \gamma_5)d \\
&\quad + \frac{1}{2} \, \overline{\nu}_e \, \gamma^{\mu}(1 + \gamma_5)\nu_e - \frac{1}{2} \, \overline{e} \, \gamma^{\mu}(1 + \gamma_5)e \\
&\quad + \cdots \\
&\quad - 2 \sin^2\theta_W \, J^{\mu}_{EM} \, ,
\end{aligned}
\tag{2.179}
$$

where the contributions of the second and third families have not been displayed explicitly. J^{μ}_{EM} and J^{μ}_{Z} are diagonal in flavor (e.g., there are no $\overline{d}\,s\,Z$ transitions). This can be traced back to the fact that all of the u^0_{mL} were assigned the same values for T^3 and Y, with a similar statement holding for the other fields. It is possible to construct alternative models in which fields of the same helicity and electric charge have different values of T^3. Such models will in general lead to flavor changing neutral current vertices unless the Yukawa couplings are constrained by additional symmetries. One of the initial motivations for the prediction of the c quark was to provide a partner for the s^0_L so that both the d^0_L and s^0_L could be placed in doublets, therefore avoiding strangeness changing neutral current vertices.

In the standard model the right-handed unitary matrices A^u_R, A^d_R, and A^e_R are unobservable, because they do not appear in the final form of the Lagrangian. (Extended models involving off diagonal Z or Yukawa couplings, right-handed charged currents, horizontal interactions, grand unified interactions, etc., may depend on the A_R.)

Some of the left-handed transformations A_L can be measured, however. The charged weak current J_W^μ can be expressed in the mass eigenstate basis as

$$\begin{aligned} J_W^\mu &= 2\bar{e}_L \gamma^\mu A_L^{e\dagger} \nu_L^0 + 2\bar{d}_L \gamma^\mu A_L^{d\dagger} A_L^u u_L \\ &= 2\bar{e}_L \gamma^\mu \nu_L + 2\bar{d}_L \gamma^\mu A_C u_L \\ &= \sum_{m=1}^{F} \bar{e}_m \gamma^\mu (1 + \gamma_5) \nu_m \\ &\quad + \sum_{mn=1}^{F} \bar{d}_m \gamma^\mu (1 + \gamma_5) A_{Cmn} u_n \ , \end{aligned} \tag{2.180}$$

where

$$A_C \equiv A_L^{d\dagger} A_L^u \tag{2.181}$$

is the generalized Cabibbo matrix, which determines the flavor structure of J_W^μ, and where

$$\nu_L \equiv A_L^{e\dagger} \nu_L^0 \ . \tag{2.182}$$

This definition is useful because the neutrinos are massless (the possibility of introducing neutrino masses is discussed in the next section). Therefore, the neutrinos cannot be distinguished from each other except by their weak interactions. We can therefore simply define ν_{mL} as the neutrino associated with e_{mL}.

From the form of J_{EM}^μ, J_Z^μ, and J_W^μ we have that

$$q_{mL} \equiv \begin{pmatrix} A_{Cmn} u_n \\ d_m \end{pmatrix}_L \tag{2.183}$$

and

$$\ell_{mL} \equiv \begin{pmatrix} \nu_m \\ e_m \end{pmatrix}_L \tag{2.184}$$

transform as SU_2 doublets while u_{mR}, d_{mR}, and e_{mR} are SU_2 singlets.

The presence of the generalized Cabibbo matrix A_C in (2.183) is due to the mismatch between the weak and mass eigenstates. For F families, A_C is an F×F unitary matrices, which can be expressed in terms of F^2 real parameters. 2F-1 of the parameters are not observable, however, because they correspond to the (unobservable) relative phases of the 2F left-handed fields. These phases can be eliminated from A_C by redefining the phases of the fermion fields. This transformation leaves the mass and Yukawa terms unchanged so long as identical phase changes are made on the right-handed fields. It should be clear that these unobservable phases correspond to the undetermined matrices K_L^u and K_L^d discussed earlier. 2F-1 of the 2F phases in K_L^u and K_L^d are usually chosen to put A_C into a convenient conventional form. The final phase does not enter the Lagrangian and is therefore arbitrary. (It should be emphasized that in a theory with more interactions, some of these phases may become observable.)

The generalized Cabibbo matrix A_C can therefore be expressed in terms of $F^2 - (2F-1) = (F-1)^2$ observable parameters. Of these, $F(F-1)/2$ correspond to Cabibbo-like rotations of the F families. The remaining $(F-1)^2 - F(F-1)/2 = (F-1)(F-2)/2$ parameters are observable phase angles. If these angles are not integer multiples of π they represent CP violation in the charged current weak interactions (the origin of which is traced back to CP violating phases in the Yukawa couplings).

For example, if there were only two families of fermions (F=2), then A_C would have one Cabibbo angle and no CP violation phases. Then

$$A_C = \begin{pmatrix} \cos\theta_C & -\sin\theta_C \\ \sin\theta_C & \cos\theta_C \end{pmatrix} , \tag{2.185}$$

so that the weak doublets are

$$\begin{pmatrix} u\cos\theta_C - c\sin\theta_C \\ d \end{pmatrix}_L$$
$$\begin{pmatrix} c\sin\theta_C + c\cos\theta_C \\ s \end{pmatrix}_L , \tag{2.186}$$

where $\theta_C \simeq 13.17°$ is the Cabibbo angle [2.39] which measures the relative strength of the $\bar{u}d$ and $\bar{u}s$ transitions. There is no CP violation in the four quark standard model.

For the case of three families (F=3) there are three rotation angles θ_i, i=1,2,3 and one CP violating phase φ. One can write

$$A_C = \begin{pmatrix} c_1 & s_1c_2 & s_1s_2 \\ -s_1c_3 & c_1c_2c_3 - s_2s_3e^{-i\delta} & c_1s_2c_3 + c_2s_3e^{-i\delta} \\ -s_1s_3 & c_1c_2s_3 + s_2c_3e^{-i\delta} & +c_1s_2s_3 - c_2c_3e^{-i\delta} \end{pmatrix} , \tag{2.187}$$

where s_i and c_i represent $\sin\theta_i$ and $\cos\theta_i$. Hence, CP violation is incorporated in the six quark version of the standard model, a result originally due to Kobayashi and Maskawa [2.58]. In this case, A_C is sometimes referred to as the KM matrix A_{KM}.

Phenomenology of the Model

The KM version of the GWS model is very successful phenomenologically. Excellent reviews have been given recently by Gaillard [2.59], by Ellis [2.60], and by Sahurai [2.61], so I will only comment on a few points.

The KM mixing matrix A_{KM} is still not well determined. Shrock and Wang [2.62] have determined $|c_1| = 0.9737 \pm 0.0025$ and $|s_3| = 0.28 {}^{+0.21}_{-0.28}$ from β decay and semi-leptonic hyperon decay (θ_1 is essentially the Cabibbo angle for small θ_3). Additional constraints can be derived: (1) from the $K_L - K_S$ mass difference [2.63-66] (in terms of the t quark mass--for $m_t = 15$ GeV it has been estimated [2.65] that $0.1 < |s_2| < 0.7$, although some of the assumptions in this estimate have been questioned [2.66]); (2) from the ε parameter which measures CP violation in the kaon system [2.63-67], yielding $s_2\, s_3 \sin\delta = O(10^{-3})$. (This test assumes that all CP violation is due to the phase φ in the KM matrix. This is true in the GWS model but not in several simple extensions. CP violation is further discussed in the next section.); and (3) from the $K_L \to \mu^+\mu^-$ rate [2.68]. Other possible constraints are reviewed in [2.59-60].

The present limits from the hyperon decays and from the $K_L - K_S$ mass difference are not so stringent as to convincingly rule out the possibility $s_2 = s_3 = 0$, which corresponds to no mixing of the t and b with the light quarks. (Alternate explanations for CP violation and b decay would then have to be invoked.) Significant improvements in the determination of A_{KM} will probably have to wait for the discovery of the t quark and for measurements of the heavy quark semi-leptonic decays.

The neutral current interaction has, of course, been discovered and has been extensively studied in neutrino-hadron, neutrino-electron, and electron-hadron interactions. The form of the interaction is compatible with (2.179), and the parameters relevant to ν-hadron scattering have been uniquely determined to lie in the region predicted by the GWS model [2.69-70]. The data are compatible with factorization, as would be expected for any model with a single Z boson [2.70-71]. If factorization is assumed the electron neutral current couplings are uniquely determined to coincide with the GWS model [2.70-71]. The entire subject has been reviewed by Kim, Langacker, Levine, and Williams [2.70], who obtain

$$\sin^2\theta_W = 0.229 \pm 0.009(\pm 0.005) \quad , \tag{2.188}$$

where the first error is from the data and the second is from the theory.

The Higgs particle η couples with strength $gm/2M_W$ to fermions. This is so small for all known fermions that the η should be extremely difficult to detect. Furthermore, the η mass is not predicted by the theory. From (2.154) we have

$$\mu_\eta = \sqrt{-2\mu^2} = \sqrt{2\lambda}\ \nu \quad , \tag{2.189}$$

where ν = 246 GeV is known from G_F; however, the quartic coupling λ is not known. Lee, Quigg, and Thacker [2.72] have shown that partial wave unitarity is satisfied at tree level only if $\lambda < 8\pi/3$, which corresponds to $\mu_\eta < [8\pi\sqrt{2}/3G_F]^{\frac{1}{2}} \simeq 1$ TeV. For larger value of λ and μ_η, the Higgs self-interactions are strong and perturbation theory breaks down. This would not necessarily be a disaster (Appelquist and Bernard [2.73], for example, have argued that the effects of a large λ on the fermion and gauge sector of the theory would be small), but 1 TeV provides a likely upper limit on μ_η. Stronger bounds can be obtained [2.74] if one requires that the effective λ be small all the way up to the Planck mass, or up to the unification mass in grand unified theories. These are of order $\mu_H < 200$ GeV.

If λ is too small, on the other hand, then corrections to the effective potential [2.29] become important. At the one loop level, (2.131) is modified to

$$V(\varphi) = \mu^2 \varphi^2 + \lambda \varphi^4 + \kappa \varphi^4 \ln(\varphi^2/v^2) \;, \qquad (2.190)$$

where $\varphi^2 \equiv \varphi^\dagger \varphi$ and κ depend on the gauge, quartic, and Yukawa couplings. Linde and Weinberg [2.75] have shown that the requirement that v be an absolute minimum (i.e., $V(v) < V(0)$) then requires $\mu_H > 6.6$ GeV. If one allows the presently observed vacuum to be metastable then lower values of μ_H are allowed. However, tunneling from the observed vacuum to the true vacuum (which would of course destroy the universe as we know it) could then occur. From the age of the universe Linde obtains [2.76] $\mu_H > 450$ MeV. Linde goes on to argue that immediately after the big bang, when the universe is incredibly hot, the vacuum would be in the symmetric phase $v = 0$ unless there is an enormous excess of leptons over baryons ($L \gtrsim 10^8$ B). In this case $\mu_H > 9.3$ GeV because otherwise the vacuum would still be in the symmetric phase. It should be emphasized that these bounds can be modified or even invalidated if heavy fermions [2.77] or additional Higgs multiplets are introduced into the theory.

It has been suggested [2.29,2.78] that μ^2 in (2.131) may be zero, in which case spontaneous symmetry breaking would still occur for sufficiently small λ. In this case, $\mu_H = 10.4$ GeV. Another possibility, suggested by supersymmetry, is $\lambda = O(g^2)$ which implies $\mu_\eta = O(M_W)$.

In the U gauge, $\mathscr{L}_{\text{ghost}}$ generates the effective interaction (2.98). This can be computed without going to a Hermitian basis by observing that $\hat{\varphi} = \eta v/\nu$ and that there are three spontaneously broken generators. Then

$$\mathscr{L}_{\text{eff}} = -3i\delta^4(0)\ell n\left[1 + \frac{\eta}{\nu}\right] \tag{2.191}$$

The GWS model is a renormalizable unitary theory of the weak and electromagnetic interactions which successfully incorporates the old Cabibbo theory. The discoveries of charm and of the neutral current strongly support the model. Such effects as nonleptonic hyperon and kaon decays, the $\Delta I = \frac{1}{2}$ rule, CP violation, the $K_L - K_S$ mass difference, and charmed hadron decays all appear to be compatible with the model, although some controversial points (especially the $\Delta I = \frac{1}{2}$ rule, CP violation, and the Cabibbo suppressed D decays) remain [2.59-60].

The model is sufficiently compelling that it is almost certainly correct at least to some degree of approximation. However, the main triumphs involve the charged and neutral current four Fermi effective interactions. The direct detection of the W and Z bosons (with masses around 80 and 91 GeV, respectively) would greatly strengthen the evidence that the underlying field theory is correct. Precision measurements of low energy processes such as $\pi_{\ell 2}$ decays or pion β decay to verify the effects of higher order weak corrections [2.79] could further establish the underlying gauge structure. Verification of the spontaneous symmetry breaking mechanism will probably require the discovery of the Higgs particle, but unfortunately this appears very difficult [2.79a] (some speculations that the Higgs particle may be replaced by a composite field will be discussed subsequently). Discovery of the t quark, studies of b and t decay, and improved experiments on CP violation could verify the detailed six quark version of the model. However, it is relatively easy to construct variations on the GWS model in which these features are modified.

2.4.4 Other Models [2.80]

An important question when one attempts to unify the weak and electromagnetic interactions with the strong interactions is whether the GWS model is the complete model of weak interactions or whether it is just an approximation to a different and probably larger underlying theory. In this section I will describe various possible extensions and variations on the model and their phenomenological implications. I will especially concentrate on those aspects which are relevant to grand unification.

The standard model can be extended by varying the number of fermion fields or their SU_2 representation assignments, by adding more Higgs fields, possibly in different representations, or by increasing the number of gauge bosons (which means going to a larger gauge group). I will discuss the implications of each of these possibilities in this section. General issues such as CP violation, the cancellation of anomalies, fermion number conservation, neutrino masses, and cosmological terms induced by spontaneous symmetry breaking will also be discussed.

Higgs Representations

The GWS model can be extended by increasing the number of Higgs fields. First consider the case that there are several complex doublets φ_m, $m = 1,2,\ldots,$ each of which contains four degrees of freedom. Only three degrees of freedom will be eliminated to give mass to the $W^\pm$ and Z, so there will be two charged and two neutral physical Higgs particles for each additional doublet (the mass eigenstates will be linear combinations of the weak eigenstates). It is possible to give these new particles very large mass [2.81], essentially by introducing large positive mass terms for the fields orthogonal to the direction of spontaneous breaking, but if they are not too massive the charged particles would be relatively easy to detect [2.82].

Unless the Yukawa couplings are restricted by additional symmetries [2.83] (e.g., by allowing only one doublet to have Yukawa couplings to each set of right-handed fermions of a given charge) the neutral Higgs fields will in general possess flavor changing Yukawa couplings. (I am using flavor in a generalized sense to include the different types of leptons as well as the different flavors of quarks.) This is because the unitary matrices A_L and A_R which make the fermion mass matrices

$$M = \sum_n \Gamma^n \, \nu_n/\sqrt{2} \tag{2.192}$$

real and diagonal do not diagonalize the individual Yukawa matrices Γ^n. The charged Higgs fields almost always connect different families (even in the presence of symmetries which keep the neutral Higgs couplings diagonal). Both the neutral and charged Higgs couplings violate parity and CP in general.

There will therefore be Higgs mediated flavor changing neutral current (FCNC) effects in most models with more than one Higgs doublet, which can contribute to such processes as $K_L \to \mu^+\mu^-$, $\mu \to e\gamma$, and to the $K_L - K_S$ mass difference. Both neutral and charged Higgs particles could also contribute to CP violating quantities such as the neutron electron dipole moment (see below). It is difficult to estimate the magnitudes of these effects because the Yukawa couplings are no longer simply proportional to the fermion masses and there are many unknown mixing angles. However, if one assumes that a typical Yukawa coupling is of order gm/M_W, that relevant mixing angles are of $O(45°)$, and that all physical Higgs masses are comparable, then one generally finds lower limits of order several hundred GeV on the Higgs masses in order to avoid conflicting with experimental limits. It should be mentioned that most of the bounds on the Higgs mass described in the last section were derived under the relatively tightly constrained hypothesis of only one Higgs doublet.

Another interesting feature of models with more Higgs doublets is that the pattern of spontaneous symmetry breaking may be more complicated. For a single Higgs doublet φ_1 one can always perform an $SU_2 \times U_1$ rotation so that

$$v_1 \equiv \langle \varphi_1 \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu_1 \end{pmatrix} , \tag{2.193}$$

with ν_1 real, so that electric charge is conserved. For two or more doublets, however, one has

$$v_n \equiv \langle \varphi_n \rangle_0 = \frac{1}{\sqrt{2}} \nu_n \begin{pmatrix} e^{i\rho_n} \sin\alpha_n \\ e^{i\sigma_n} \cos\alpha_n \end{pmatrix} \tag{2.194}$$

where α_1, σ_1, ρ_1 can be chosen to be zero by convention but the other angles are determined by the potential. If any of the α_n are non-zero, $SU_2 \times U_1$ will be completely broken and there will be no conserved electric charge. Non-trivial values of ρ_n or σ_n are usually associated with CP violation.

For the case of two doublets, for example, the potential is

$$V(\varphi_1, \varphi_2) = V_A(\varphi_1^\dagger \varphi_1, \varphi_2^\dagger \varphi_2) + V_B(\varphi_1, \varphi_2) , \tag{2.195}$$

where

$$V_A = +\mu_1^2 \varphi_1^\dagger \varphi_1 + \mu_2^2 \varphi_2^\dagger \varphi_2 + \lambda_1(\varphi_1^\dagger \varphi_1)^2 + \lambda_2(\varphi_2^\dagger \varphi_2)^2 + \lambda_3 \varphi_1^\dagger \varphi_1 \varphi_2^\dagger \varphi_2 \tag{2.196}$$

and

$$V_B = \lambda_4[(\varphi_1^\dagger \varphi_2)^2 + (\varphi_2^\dagger \varphi_1)^2] + \lambda_5 \varphi_1^\dagger \varphi_2 \varphi_2^\dagger \varphi_1 \ . \tag{2.197}$$

I have simplified the problem by assuming the λ's are real and by imposing invariance under the discrete symmetry $\varphi_1 \to -\varphi_1$, $\varphi_2 \to +\varphi_2$ on V. In terms of the VEV's, V_A depends only on the magnitudes ν_1^2 and ν_2^2, while

$$V_B = \frac{1}{4} \nu_1^2 \nu_2^2 \cos^2\alpha_2[2(\cos 2\sigma_2)\lambda_4 + \lambda_5] \ . \tag{2.198}$$

Assuming that V_A is such that ν_1 and ν_2 are non-zero, the potential will be minimized for $\cos\alpha_2 = 0$ if $\lambda_5 > 2|\lambda_4|$ and for $|\cos\alpha_2| = 1$ if $\lambda_5 < 2|\lambda_4|$. These two cases correspond to the nonconservation and conservation of electric charge, respectively. In the latter case the sign of $\cos 2\sigma_2$ will be opposite of that of λ_4. The implications for CP violation will be discussed later.

In general, the pattern of symmetry breaking depends in a complicated way on the values of the parameters in the Higgs potential. Both the relative orientations and the number of fields with non-zero VEV's can be changed by changing the parameters. (The couplings must be chosen so that the total potential is bounded from below.)

One can also introduce Higgs fields belonging to different SU_2 representations. If such fields have a non-zero VEV they will violate the relation $M_Z^2 = M_W^2/\cos^2\theta_W$, which holds for any number of Higgs doublets (and singlets). Instead, one has (assuming electric charge conservation)

$$M_Z^2 = (g^2 + g'^2) \sum_n (t_n^3)^2 \nu_n^2$$

$$M_W^2 = \frac{1}{2} g^2 \sum_n [t_n(t_n + 1) - (t_n^3)^2] \nu_n^2 \tag{2.199}$$

$$\rho \equiv \frac{M_W^2}{\cos^2\theta_W M_Z^2} = \frac{\sum_n [t_n(t_n + 1) - (t_n^3)^2] \nu_n^2}{2 \sum_n (t_n^3)^2 \nu_n^2} ,$$

where ν_n is the VEV of a Higgs field with the values t_n and t_n^3 for T and T^3. The effective neutral current interaction then becomes

$$\mathscr{L}_Z^{eff} = \frac{G_F}{\sqrt{2}} \rho J_Z^\mu J_{\mu Z} . \tag{2.200}$$

From fits to the neutral current data with ρ allowed to be a free parameter, one has [2.70]

$$\begin{aligned} \rho &= 0.992 \pm 0.017 \ (\pm 0.011) \\ \sin^2\theta_W &= 0.224 \pm 0.015 \ (\pm 0.012) . \end{aligned} \tag{2.201}$$

Hence, ρ is very close to the value expected if there are only Higgs doublets.

Fermion number violating Yukawa couplings can be written if SU_2 Higgs singlets [2.84] or triplets [2.85] are introduced, as will be discussed subsequently.

The Induced Cosmological Term

One very serious difficulty with the idea of spontaneous symmetry breaking concerns the non-zero value of the potential when it is evaluated at the minimum $\varphi = v$ (2.153). As pointed out by Zel'dovich [2.86] such a vacuum self-energy term must be interpreted as a cosmological constant

$$\Lambda = \frac{8\pi G_N}{c^4} V(v) \tag{2.202}$$

in the Einstein equations, where G_N is the gravitational constant.

For the GWS model, we have

$$\Lambda = \frac{2\pi G_N}{c^4} \mu^2 \nu^2 = -\left(\frac{\pi G_N}{\sqrt{2}\, G_F}\right) \mu_\eta^2 = -1.3 \times 10^{-33} \mu_\eta^2 \tag{2.203}$$

However, according to the observed limits on the deceleration of the expansion of the universe [2.87],

$$|\Lambda_{obs}| < 10^{-56} \text{ cm}^{-2} \approx 4 \times 10^{-84} \text{ GeV}^2 . \tag{2.204}$$

Hence, for $\mu_\eta > 6.6$ GeV (Section 2.43) $|\Lambda|$ is predicted to be fifty-two orders of magnitude too large. (This result was first pointed out by Linde [2.88] and Veltman [2.89].) Dreitlein [2.90] interpreted (2.203) and (2.204) as a limit of 5.5×10^{-26} GeV on the Higgs mass (he actually used a somewhat different limit on Λ), but Veltman argued [2.91] that such a light particle is ruled out because it would mediate a macroscopic long range force seven orders of magnitude stronger than gravity. Also, such a small value appears to be highly unlikely when one considers higher order corrections to V (2.190).

The whole problem can be eliminated by replacing $V(\varphi)$ by

$$V'(\varphi) = V(\varphi) - V(v) , \tag{2.205}$$

so that $V'(v) = 0$. The addition of a constant vacuum self-energy density to $\mathscr{L}$ does not affect particle physics. It can be interpreted as a positive primordial cosmological term that cancels (to better than one part in 10^{52}!) the cosmological constant induced by spontaneous symmetry breaking. One possible danger in such a scheme is that early in the history of the universe when the temperature T was large compared to $M_{W^\pm, Z}$ the vacuum may have been in a different phase with $\langle \varphi \rangle = 0$ (temperature dependent phase transitions are discussed in Chapter 6). At that time, the Λ induced by SSB would go away but the large primordial Λ would remain. As pointed out by Bludman and Ruderman [2.87], however, the vacuum energy density

$$\rho_{vac}(T) = -V(v) = \frac{+\mu^4}{4\lambda} \tag{2.206}$$

will be small compared to the thermal energy density [2.92]

$$\rho_{thermal}(T) = 0.33\ g_t\ T^4 \tag{2.207}$$

($g_t = g_B + 7g_F/8$, where g_B and g_F are the number of boson and fermion degrees of freedom that are light compared to T) for $T >> T_c$, where T_c is the critical temperature above which $\langle \varphi \rangle = 0$. ($T_c$ is of the order [2.87,2.93] of the smaller of μ_H/g and M_W/g.) Hence, the primordial cosmological term would not be important at early times ($T >> T_c$), although for a small Higgs mass it could be of some relevance [2.93] at the time for which $T \gtrsim T_c$. Similar statements apply to a possible phase transition which restores the symmetry in a grand unified theory.

A much more serious objection to adding a constant to V is that the primordial cosmological term and the constant induced by spontaneous symmetry breaking appear to be unrelated. There is no known reason for these two terms to cancel each other to one part in 10^{52}. (Note that the cancellation must be maintained in the presence of radiative corrections to the Higgs potential.) This unnatural cancellation is probably the most serious difficulty with all models of spontaneous symmetry breaking, but it is only the first of several delicate adjustments of parameters that occur in the standard model and in most grand unified theories.

The possibility of eliminating elementary Higgs fields will be discussed in Section 2.5.3. The restoration of symmetries at high temperature is described in Chapter 6.

Anomalies

When one modifies the fermion or gauge structure of the theory it is important to avoid the introduction of anomalies [2.94-95], which are singularities associated wtih the fermion triangle diagram contributions to the vertex of three currents, as shown in Fig. 2.8. If one or three of the vertices involves an axial vector coupling (γ_5) the diagram diverges linearly. This linear divergence leads to an anomalous divergence of the currents in perturbation theory that is not revealed by formal manipulation of the field equations. If one or more of the currents is associated with a global symmetry of the theory, the anomalous divergence does not cause any particular problems, and it can even be useful [2.94]. If the currents are all associated with gauge symmetries, however, then the diagram contributes to the vertex of the three gauge fields that couple to the currents. In this case, the vertex cannot be regularized in a way consistent with the gauge invariance of the theory. Gauge invariance and renormalizability are therefore destroyed.

The anomaly coefficient A_{ijk} in the vertex of currents i, j, and k is independent of the fermion masses. It is [2.95]

$$A_{ijk} = \mathrm{Tr}\, \gamma_5\, L^i \{L^j, L^k\} \ , \qquad (2.208)$$

where $L^i = L^i_L P_L + L^i_R P_R$ is the fermion representation matrix. The trace extends over both Dirac and group indices. The Dirac part can be carried out to give

$$A_{ijk} \simeq 2\mathrm{Tr}\, L^i_L \{L^j_L, L^k_L\} - 2\mathrm{Tr}\, L^i_R \{L^j_R, L^k_R\} \equiv 2(A^L_{ijk} - A^R_{ijk}) \qquad (2.209)$$

In the GWS model the quark and lepton contributions to the anomalies cancel so that $A^L = A^R = 0$. (Except for the B^3 vertex, for which $A^{L,R} = 2\mathrm{Tr}\, Y^3_{L,R} = 2F\, \Sigma\, y^3_{L,R}$. Then $A^L = A^R = -4F/9$. One must remember that each quark doublet comes in three colors.) In any non-chiral model ($L_L = L_R$) we have $A^L = A^R$ so that the anomalies are zero. More generally, the demand that $A_{ijk} = 0$ is a useful constraint on the construction of gauge theories.

Fermions

Let us now consider the modification of the fermions in the six quark GWS model. The simplest modification is to increase the number of families F. However, the primordial helium abundance in the universe is predicted to grow rapidly with the number F_ν of massless (or nearly massless) neutrinos. The observed abundance strongly suggests [2.96] that $F_\nu = 3$. This limit could be circumvented if there were an enormous neutrino degeneracy in the universe [2.97], but this appears unlikely if grand unified theories are correct [2.98]. (The various cosmological constraints on massive stable or unstable neutral leptons, superweakly coupled particles such as right-handed neutrinos, and other exotic objects are reviewed by Steigman [2.92].) Possible accelerator limits have also been suggested [2.99].

There are no firm upper limits on the number F_e of charged leptons or on the number F_q of quark doublets. If one assumes that each charged lepton is associated with a massless neutrino then $F_e = F_\nu$. However, one can easily get around this constraint by adding new lepton families

$$\begin{pmatrix} N_m^0 \\ e_m^0 \end{pmatrix}_L \qquad N_{mR}^0 \,, \tag{2.210}$$

where $N_{mL,R}^0$ is a heavy neutral lepton.

For the quarks the strongest constraint comes from the cancellation of anomalies, which requires $F_q = F_e$ (again one can get around this, e.g., by putting both the left and right handed quarks in doublets). A weaker limit of $F_q \leq 8$ can be derived if one demands asymptotic freedom for the strong interactions. (For $F_q > 8$ the strong coupling constant would grow with Q^2, but this behavior would not begin until $|Q^2|$ is greater than the mass^2 of the heavy quarks. See Section 2.5.2.) Maiani et al. [2.74] have also shown that the requirement that all gauge couplings in the standard model be finite below the Planck mass leads to $F_q \leq 8$ and also that the masses m_n of any new fermions be less than 160 GeV.

There are other arguments that the masses m_n of new fermions should not be too large. Hung and Politzer and Wolfrom [2.77] have shown that for the effective potential (2.190) to be bounded from below (i.e., for $\kappa > 0$) requires $[\Sigma\, m_n^4]^{\frac{1}{4}} < 873$ GeV if tree unitarity is demanded ($\lambda < 8\pi/3$). This bound could be evaded by allowing more Higgs fields or (possibly) by including two loop terms in V. Finally, heavy fermion loops renormalize the value of ρ in (2.199) [2.100]. If one assumes that $\rho = 1$ initially (only Higgs doublets) then the upper limit (2.201) on ρ implies $m < 500$ GeV for a charged lepton with a massless partner [2.70]. Limits of the same order apply to the mass difference of new quarks in a doublet. These various bounds on fermion masses suggest that the number of families is limited.

The representation assignments of the fermions can be altered, by placing some right-handed fermions in doublets, for example. Especially attractive are the vectorlike theories [2.101], in which the left- and right-handed fermion representation matrices (in the weak basis) are equal ($L^i_L = L^i_R$). Such theories are reflection invariant, at least in the gauge sector (parity can be spontaneously broken), and anomalies are automatically absent. The simplest possibility, in which all fermions are placed in doublets [2.102], is now ruled out by the observed parity violation in the neutral current interactions. Vectorlike models involving singlets and doublets, such as

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \begin{pmatrix} U \\ D \end{pmatrix}_R \quad u_R,\ d_R,\ U_L,\ D_L \tag{2.211}$$

for one family (u, d) of light quarks and one (U, D) of heavy quarks, cannot be ruled out but are not required experimentally.

The amount of mixing between doublets and singlets in such models (whether or not they are vectorlike) is restricted by the neutral current data. The present limits for mixing between right handed singlets and doublets are [2.70]

$$\begin{aligned} \sin^2\alpha_u &\leq 0.103 \\ \sin^2\alpha_d &\leq 0.348 \\ \sin^2\alpha_e &\leq 0.064\ , \end{aligned} \tag{2.212}$$

where $u_R \sin\alpha_u$ and $u_R \cos\alpha_u$ are the components of u_R in a doublet and singlet, respectively. Even stronger constraints, both from charged and neutral currents, hold for the left-handed fields. Finally, the absence of FCNC effects for the light fermions requires that the mixing between singlets and doublets must either be extremely tiny or restricted in form [2.103]. (Mixing can be eliminated entirely by the introduction of appropriate symmetries.)

There have been several models in which the b quark cannot decay via ordinary charged current weak effects, either because extra symmetries prevent the $(t, b)_L$ doublet from mixing with the other quarks [2.104] or because the t quark does not exist and the b is in an SU_2 (unmixed) singlet [2.105]. In such models extra interactions involving the exchange of Higgs bosons or additional gauge bosons are needed to mediate b decay (which is always semi-leptonic in the models cited) and CP violation.

Global Symmetries and Neutrino Masses

The GWS model possesses several global U_1 symmetries as described in Section 2.3.1. The corresponding quantum numbers are quark number N_q ($N_q = 3B$ where B is baryon number) and individually conserved electron, muon, and τ numbers N_e, N_μ, and N_τ. The total lepton number L and the fermion number $F \equiv N_q + L$ are therefore also conserved. Strangeness, charm, etc. are violated by the charged current weak interactions. The origin of these violations can be traced back to the off diagonal Yukawa interactions.

The separate conservation of N_e, N_μ, and N_τ in the GWS model is due to the assumed masslessness of the neutrinos. However, the experimental limits on the neutrino masses are not very stringent. The best laboratory limits are $m_{\nu_e} < 35$ eV [2.106], $m_{\nu_\mu} < 0.57$ MeV [2.107], and $m_{\nu_\tau} < 250$ MeV [2.108]. Much stronger limits come from the cosmological requirement that the neutrino contribution to the energy density of the universe must not exceed the observed density [2.92]. It is $\Sigma\, m_{\nu_n} < 50$ eV, where the sum extends over all light stable neutrinos. Recently, Schramm and Steigman [2.109] have advocated massive neutrinos. They argue that if the most massive neutrino has a mass in the range 3 eV $< m_\nu <$ 10 eV, it could account for the missing mass in galactic clusters without violating Tremaine and Gunn's [2.110] bound $m_\nu <$ 10 eV (to avoid contributing too much mass to binary galaxies). Witten [2.111] has given similar arguments, but prefers larger masses (tens of eV), which could close the universe.

One can give mass to the neutrinos in the GWS model simply by introducing right-handed neutrino fields ν^0_{nR} which couple to the lepton doublets with Yukawa couplings analogous to (2.142). This will lead to a neutrino mass matrix and an observable leptonic mixing matrix $A_{lep} \equiv A_L^{e\dagger} A_L^{\nu}$ analogous to the Cabibbo matrix for the quarks [2.112-113]. This would have several interesting consequences. Kolb and Goldman [2.112] have recently discussed the possibility that the ν_τ is massive ($m_{\nu_\tau} >$ 10 MeV) and unstable. Assuming that A_{lep} is non-diagonal, then N_e, N_μ, and N_τ will no longer be separately conserved, although the total lepton number L would be. A dramatic consequence would be the possibility of observing neutrino oscillations [2.114-115]. These should be detectable in laboratory experiments if typical neutrino mass differences are of order of a few eV, assuming reasonably large mixing angles [2.115]. For mass differences larger than $\simeq 10^{-6}$ eV, electron neutrinos could oscillate sufficiently rapidly to help explain the missing solar neutrinos [2.116].

Additional Higgs particles, additional gauge bosons, and the mixing of heavy neutral leptons as in (2.110) could also lead to the violation of N_e,

N_μ, and N_τ and could cause such FCNC processes as $\mu \to e\gamma$ (the amplitude for the latter from the mixing of light neutrinos is non-zero but negligibly small [2.117]).

Majorana and Dirac Masses

I have so far assumed that the total lepton and fermion numbers are absolutely conserved. This is guaranteed in the GWS model, but need not hold if right-handed neutrino fields or certain types of additional Higgs fields are introduced. In particular, the possibility of fermion number violating Majorana mass terms for the neutrinos becomes possible [2.118]. I will discuss this issue in some detail because many grand unified models predict such effects. For a more detailed treatment, see [2.119].

First consider the case of one family: $(\nu e)_L$, ν_R, e_R. In addition to the Dirac mass term $M_D \bar{\nu}_L \nu_R$ + H.C., generated by the Higgs doublets, one can introduce a Majorana mass term

$$m_M \overline{\nu_L^c} \nu_R + m_M^* \bar{\nu}_R \nu_L^c = m_M \nu_R^T C \nu_R + m_M^* \nu_L^{cT} C \nu_L^c , \tag{2.213}$$

where C is the charge conjugation matrix introduced in (2.110). This term can be introduced into the Lagrangian as a bare term since ν_R is a singlet under $SU_2 \times U_1$, or from the Yukawa couplings to a singlet Higgs field. Recall that

$$\nu_{L,R} \equiv P_{L,R} \nu ,$$

where ν is a four component field, annihilate L and R neutrinos and that

$$\nu_{L,R}^c \equiv P_{L,R} \nu^c = P_{L,R} C \bar{\nu}^T = C \bar{\nu}_{R,L}^T \tag{2.214}$$

annihilate L and R anti-neutrinos. ν_L and ν_R^c are members of doublets, while ν_R and ν_L^c are singlets. The Majorana mass term violates lepton (and fermion) number by two units.

m_M and m_D can be taken to be real, by an appropriate phase change on the fields, so that the overall mass term is

$$m_D(\bar{\nu}_L \nu_R + \bar{\nu}_R \nu_L) + m_M(\overline{\nu_L^c} \nu_R + \bar{\nu}_R \nu_L^c) = m_D \bar{\nu}\nu + m_M \bar{\eta}\eta \qquad (2.215)$$

where

$$\begin{aligned} \nu &= \nu_L + \nu_R \\ \eta &= \nu_L^c + \nu_R \end{aligned} \qquad (2.216)$$

In the familiar case that $m_M = 0$, we identify ν as the mass eigenstate. If $m_D = 0$, on the other hand, the new field η is the mass eigenstate. η is a Majorana or self conjugate field in the sense that

$$\eta \underset{\hat{C}}{\rightarrow} \eta^{\hat{c}} \equiv C(\bar{\eta})^T = \eta \ . \qquad (2.217)$$

Note, however, that $\hat{C}$ is not the same charge conjugation as defined in (2.214), under which

$$\eta \underset{C}{\rightarrow} \eta^c = \nu_L + \nu_R^c \neq \eta \qquad (2.218)$$

In the general case with $m_D \neq 0$, $m_M \neq 0$ one must regard (ν_L, ν_L^c) and (ν_R^c, ν_R) as pairs of independent left- and right-handed fields. Linear combinations with definite mass (but not definite lepton number) are chosen to diagonalize the mass matrix. The most general mass matrix (for one family) is

$$(\bar{\nu}_L \ \overline{\nu_L^c}) \begin{pmatrix} a & d \\ d & s \end{pmatrix} \begin{pmatrix} \nu_R^c \\ \nu_R \end{pmatrix} + \text{H.C.} \qquad (2.219)$$

where $d = m_D/2$ (the off diagonal entries are equal because $\bar{\nu}_L \nu_R = \overline{\nu}^c_L \nu^c_R$) and $s = m_M$. The other Majorana term a is forbidden by the gauge symmetry unless a Higgs triplet $\vec{\rho} = (\rho^o, \rho^-, \rho^{--})^T$ is introduced into the model. Then, the Yukawa coupling

$$(\bar{\nu}_L\ \bar{e}_L)\ \vec{\tau}\cdot\vec{\rho} \begin{pmatrix} e^c_R \\ -\nu^c_R \end{pmatrix} = (\bar{\nu}_L\ \bar{e}_L) \begin{pmatrix} \rho^- & \sqrt{2}\,\rho^o \\ \sqrt{2}\,\rho^{--} & -\rho^- \end{pmatrix} \begin{pmatrix} e^c_R \\ -\nu^c_R \end{pmatrix} \tag{2.220}$$

will generate an a term and violate L by two units if $\langle\rho^o\rangle_0 \neq 0$. (For $\langle\rho^o\rangle_0 = 0$, L could be conserved if $\vec{\rho}$ were assigned two units of "lepton number".)

The mass matrix in (2.220) can be diagonalized by separate left and right-handed unitary transformations (these are complex conjugates of each other up to a matrix of phases [2.119]), as in (2.172), to give

$$m_1\,\overline{\nu}^c_{1L}\,\nu_{1R} + m_2\,\overline{\nu}^c_{2L}\,\nu_{2R} + \text{H.C.} = m_1\,\bar{\nu}_1\,\nu_1 + m_2\,\bar{\nu}_2\,\nu_2\ , \tag{2.221}$$

where ν^c_{nL} and ν_{nR} are linear combinations of ν_L and ν^c_L and of ν_R and ν^c_R, respectively, and $\nu_n = \nu^c_{nL} + \nu_{nR}$ are the mass eigenstates. ν_1 and ν_2 are Majorana fields. In the special case $m_1 = m_2$ they combine to form a single Dirac neutrino [2.119].

In the case of F families, the mass term is

$$(\bar{\nu}^o_1 \ldots \bar{\nu}^o_F\ \overline{\nu}^{oc}_1 \ldots \overline{\nu}^{oc}_F)_L\ m \begin{pmatrix} \nu^{oc}_1 \\ \vdots \\ \nu^{oc}_F \\ \nu^o_1 \\ \vdots \\ \nu^o_F \end{pmatrix}_R + \text{H.C.}\ , \tag{2.222}$$

where m is a $2F \times 2F$ matrix of the form

$$m = \left(\begin{array}{c|c} A & D \\ \hline D^T & S \end{array}\right) , \tag{2.223}$$

where D is an $F \times F$ Dirac mass matrix, and S and A are symmetric $F \times F$ Majorana mass matrices generated by Higgs singlets (or bare mass terms) and triplets, respectively. One can then find 2F mass eigenstates ν^c_{nL} (which are linear combinations of the ν^o_L and ν^{oc}_L), and 2F eigenstates ν_{nR} by diagonalizing m. The fields $\nu_n \equiv \nu^c_{nL} + \nu_{nR}$ describe 2F Majorana particles.

The laboratory and cosmological limits on neutrino masses and the approximate validity of Cabibbo universality require that any very massive physical neutrino states should be approximately orthogonal to the weak eigenstates ν^o_{nL}. Furthermore, a Majorana mass term a $\bar{\nu}^o_L \nu^{oc}_R$ would lead to neutrinoless double β decay, $(Z) \to (Z + 2)e^-e^-$, as shown in Fig. 2.9 (a is treated as a perturbation). One can adopt the estimates of Halprin et al. [2.120] to determine that a < 1 KeV to avoid conflict with the experimental limits [2.121]. The constraints from the limits on the physical neutrino masses will generally lead to a more stringent restriction on a. Right-handed Majorana mass terms s $\bar{\nu}^{oc}_L \nu^o_R$ are not strongly restricted because the ν^o_R interacts only by Yukawa interactions and mass terms. If these are sufficiently weak [2.92] the cosmological mass bounds do not apply.

If neutrino masses are allowed, one has to worry about why they are so much smaller than other fermion masses. One interesting possibility, suggested by Gell-Mann, Ramond, and Slansky [2.122] is that a = 0 (no Higgs triplets) and s >>> m_D, where m_D is of the same order of magnitude as other fermion masses. A large s could possibly be generated by whatever mechanism breaks a grand unified theory down to $SU_2 \times U_1$. Then the mass eigenstates are very nearly $\nu_L^{oc} + \nu_R^o$ and $\nu_L^o + \nu_R^{oc}$ with masses $\simeq s$ and $m_D^2/4s \lll m_D$, respectively. For $s \simeq 10^{14}$ GeV and $m_D/2 \simeq 1$ GeV, the light neutrino mass will then be $\simeq 10^{-5}$ eV, in the range relevant for the solar neutrino problem. There will be an induced Majorana mass term $a\, \bar{\nu}_L^o\, \nu_R^{oc}$ as shown in Fig. 2.9, with $a \sim m_D^2/4s$, but this is much too small to lead to observable neutrinoless double β decay. Other schemes suggested by various grand unified theories are described in Chapter 6.

Another possibility is that all of the Majorana and Dirac mass terms in (2.223) are small and of the same order of magnitude. In this case the mass eigenstates will contain significant admixtures of the weak doublets ν_{nL}^o and of the weak singlets ν_{nL}^{oc}. Oscillations could then occur not only between the doublets but between the doublets and singlets [2.119].

Additional Gauge Bosons

The GWS model can also be extended by adding additional gauge bosons (i.e., by going to a larger gauge group). One possibility is to consider the group $G \times U_1$, where SU_2 is a subgroup of G. Various $SU_3 \times U_1$ models, for example, have been considered [2.123], primarily for obsolete reasons (e.g., high y anomaly, lack of atomic parity violation, anomalous neutrino induced trimuon events). One can easily generate vectorlike models with the fermions in three dimensional representations (which contain an SU_2 doublet and singlet), but the versions still compatible with the charged and neutral current data generally require a proliferation of heavy fermions and make few interesting predictions for the light particles. FCNC are also a problem in such models.

Another possibility is to consider direct products $SU_2 \times U_1 \times G$. In particular, much attention was devoted [2.124], again mainly for obsolete reasons (e.g., absence of atomic parity violation), to the group $SU_{2L} \times SU_{2R} \times U_1'$, where the subscripts mean that the left handed fermions transform as doublets and singlets with respect to SU_{2L} and SU_{2R}, respectively. The reverse holds for the right-handed fermions. The Langrangian for such a model can be made left-right symmetric (i.e., symmetric under the interchange of left and right labels (parity)) by the imposition of a discrete symmetry. Except for mixing effects, the $W_R^{\pm}(W_L^{\pm})$ bosons of the $SU_{2R}(SU_{2L})$ group couple only to right (left) handed currents. Parity could be spontaneously broken to make the $W_R^{\pm}$ much more massive than the $W_L^{\pm}$. Depending on the details of the symmetry breaking the neutral current interactions may or may not coincide with the GWS model. In the former case, the B boson of the $SU_{2L} \times U_1$ subgroup is a linear combination of the W_R^0 and the B' associated with U_1'. $SU_{2L} \times SU_{2R} \times U_1'$ models require right-handed neutrinos, so in general the neutrinos will be massive.

Another popular extension in this class is to the group $SU_2 \times U_1 \times U_1'$ [2.125]. The extra U_1' factor may remain from the breakdown of a larger group, such as the models described above.

Horizontal Symmetries [2.126]

Finally, there have been many models in which $SU_2 \times U_1$ is extended to $SU_2 \times U_1 \times G_H$, where G_H represents a horizontal group. Horizontal means that the G_H transformations act on the family indices; they rotate or interchange one family into another or at least distinguish between the families (e.g., if

G_H is a U_1 group with different charges for the different weak basis families). The name horizontal comes from the practice of displaying the SU_2 doublets side by side. The $SU_2 \times U_1$ interactions are then referred to as vertical interactions. Much of the early motivation for imposing horizontal symmetries was to find relations between fermion masses and mixing angles (e.g., to predict the t quark mass), but more generally the existence of several families suggests the existence of symmetries to relate them. Many of the horizontal groups that have been proposed have involved discrete permutation groups (which could conceivably be the remnant of a broken continuous group). G_H can also be a continuous group, in which case it is most likely a gauge symmetry. Otherwise, there would be unwanted Goldstone bosons associated with the spontaneous breaking of G_H. (An alternative would be to give masses to these bosons by softly breaking G_H in the dimension 2 or 3 terms in $\mathcal{L}$.) Horizontal gauge bosons can typically mediate FCNC processes. This is usually not intrinsic to the horizontal interaction because the bosons simply transmit flavor from one side of a diagram to the other. However, the mixing between fermions and/or the mixing of horizontal with vertical bosons will generally lead to FCNC effects such as $\mu \to e\gamma$, $K_L \to \mu^+\mu^-$, and a contribution to the $K_L - K_S$ mass difference. The reader is referred to [2.126] for a discussion of the mass and mixing angle predictions of horizontal symmetries (which have not been especially successful). For three families an obvious choice for a horizontal gauge group to relate them would be SU_3, with the three families forming a triplet. However, this theory would have anomalies unless right-handed neutrinos are added [2.127]. Wilczek and Zee [2.127] have therefore considered a horizontal SU_2 with the three families in a triplet. Higgs bosons contributing to the fermion mass matrix must be singlets, vectors, or tensors under the horizontal group (and doublets under the ordinary SU_2). Gell-Mann, Ramond, and Slansky [2.122] have emphasized that the approximate mass matrices

$$M^{u,d,e} \simeq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c^{u,d,e} \end{pmatrix}, \tag{2.224}$$

which could explain why the third family is so massive compared to the first two, would require a very unnatural collaboration between a Higgs singlet and tensor.

Masses of Additional Bosons

There is no experimental evidence for any of these extensions, and the effects of the additional interactions on the light fermions must be suppressed, generally by requiring the new bosons to be very massive and/or by making their gauge couplings weak and/or by cleverly arranging the representations and mixings of the fermions. Assuming all gauge couplings to be comparable, which is plausible if all of the groups are ultimately embedded in an underlying grand unified theory, one typically finds that right-handed charged bosons or additional neutral bosons with flavor conserving interactions must be at least five or ten times as massive as the $W_L^{\pm}$ and Z [2.128]. Limits on horizontal gauge boson masses are more stringent because they can mediate FCNC processes. Their masses should be $\simeq 10^3$-10^4 times those of the $W^{\pm}$ and Z unless the relevant vertices are suppressed by small mixing angles or the gauge couplings are very weak [2.126-127].

Nevertheless, such models are interesting to consider because they often have attractive theoretical features (depending on the taste of the theorist!), and they may be relevant to CP violation; FCNC processes; the spectrum, production, and decay of heavy fermions, etc. Finally, many grand unified theories lead to weak interaction subgroups larger than $SU_2 \times U_1$.

CP Violation

In the GWS model with a single Higgs doublet the only source of CP violation (other than the nonperturbative effects discussed below) are in the couplings of the charged W bosons to fermions. In the four quark model there is no CP violation, while in the six quark model there is one CP violating phase δ (2.187). Observable CP violation is associated only with phase differences between amplitudes, so it is convenient to choose the conventional form of A_C so that δ only appears in transitions involving at least one heavy (c, b, or t) quark. CP violation in the kaon system is primarily associated with phases in the $\Delta S = 2$ effective interaction

$$H_{eff}^{\Delta S=2} = c\ \bar{s}_L \gamma_\mu d_L \bar{s}_L \gamma^\mu d_L + \text{H.C.} \tag{2.225}$$

The phases in c are generated by box diagrams in which two charged W bosons are exchanged [2.67].

Generalizations of the GWS model allow many sources of CP violation. These include the exchange of additional charged gauge bosons (such as the $W_R^{\pm}$ of $SU_{2L} \times SU_{2R} \times U_1$ models [2.129]) and of charged Higgs particles [2.130] in models with more than a single Higgs doublet. Models in which CP violation is associated with the exchange of charged particles are typically in the milliweak to microweak range which means that the strength of the CP violation in $\Delta S = 1$ processes is of order 10^{-3} to 10^{-6}. There are also models in which CP violation is mediated by the exchange of a neutral boson (a heavy gauge boson

Z' associated with a horizontal interaction [2.131], or a neutral Higgs particle [2.132]) with off diagonal couplings. These models can contribute to the $\Delta S = 2$ effective Hamiltonian at tree level, so it is necessary for the interactions to be superweak (i.e., the CP violating part of the $\Delta S = 1$ interaction is of order 10^{-9}) either by making the boson extremely massive or by choosing very small mixing angles or coupling constants. The latter possibility is unlikely for a neutral gauge boson if the new interaction is to be combined with the ordinary interactions in a grand unified theory. Of course, it is possible to have several distinct sources of CP violation all in the same model.

All of the models are arranged to approximately reproduce the results of the superweak model for the $K_L \to 2\pi$ decays. Possible experimental tests to distinguish the models include: (a) a precise measurement of the ε' parameter of $K_L \to 2\pi$ decays, which is sensitive to the phase difference between the $\Delta I = 3/2$ and $\Delta I = 1/2$, $\Delta S = 1$ amplitudes. The KM model and the models with charged Higgs particles tend to give [2.133,2.130] $|\varepsilon'/\varepsilon| \leq 0.02$. The simplest $W_R^{\pm}$ model gives [2.129] $\varepsilon' \simeq 0$, as do the superweak models [2.131-132]. (b) A measurement of $r \equiv \eta_{+-0}/\eta_{+-}$, where η_{+-0} is $A(K_S \to \pi^+\pi^-\pi^0)/A(K_L \to \pi^+\pi^-\pi^0)$. The GWS model [2.134] and the superweak models give $r \simeq 1$ (i.e., the $K_S \to 3\pi$ decay is due to $K_1 - K_2$ mixing effects), while the $W_R^{\pm}$ models give [2.129] r different from unity. In general, soft pion arguments can be used to relate the $K_S \to 3\pi$ and $K_L \to 2\pi$ amplitudes [2.134-135], with the result that r can differ significantly from unity only in milliweak models in which the $\Delta S = 1$ amplitude contains two or more pieces with different helicity structures, such as models involving charged Higgs or W_R bosons in addition to $W_L^{\pm}$. However, the simple charged Higgs model of Ref. [2.130] gives [2.135] $r \simeq 1$. (c) The weak interaction contribution to the electric dipole moment of the neutron. The models involving charged Higgs particles typically yield [2.130] $d_N \simeq 10^{-25}$ e-cm, whereas the other models typically give $d_N \simeq 10^{-30}$ e-cm. Nonperturbative contributions to d_N are discussed below. (d) CP violating effects in the decays of heavy quarks. For further discussion of the phenomenology of CP violation, see [2.136].

A serious complication is that CP violation can be associated with nonperturbative effects in the strong interaction (QCD) sector of the theory. As will be described in the next section, the normal QCD Lagrangian is CP invariant. However, it is possible to add an additional term

$$\mathscr{L}_\theta = \frac{\theta}{64\pi^2} \varepsilon^{\mu\nu\sigma\tau} F^i_{\mu\nu} F^i_{\sigma\tau} , \tag{2.226}$$

where $F^i_{\mu\nu}$ are the gauge covariant field tensors of QCD, to the Lagrangian. $\mathscr{L}_\theta$ violates both P and T. $\mathscr{L}_\theta$ can be written as a total divergence, but nevertheless the contribution of $\mathscr{L}_\theta$ to the action does not vanish because nonperturbative instanton effects [2.137-138] generate non-zero surface terms. Because of anomalies $\mathscr{L}_\theta$ is generally not invariant under the global unitary transformations that are needed to make the quark mass matrix real and diagonal [2.139]. When expressed in this basis, the effective θ in (2.226) is

$$\theta = \theta_{QCD} + \theta_{QFD} , \tag{2.227}$$

where θ_{QCD} is the bare value from the QCD Lagrangian and

$$\theta_{QFD} = \arg \det m_L , \tag{2.228}$$

where $\det m_L = (\det m^u_L)(\det m^d_L)$ is the determinant of the quark mass matrix, with $m_{L,R}$ defined in (2.45).

Therefore, CP can be violated not only by the explicit weak interaction effects described above but also by $\mathscr{L}_\theta$. Note that $\theta \neq 0$ in general even if $\theta_{QCD} = \bar{0}$. $\mathscr{L}_\theta$ is expected to contribute $d_N \simeq c\theta/m_N$ to the neutron electric dipole moment. From the limit [2.140] on d_N one typically requires $\theta < 10^{-8}$-10^{-10}, depending on c. In order to see the effects of this constraint, it is useful to divide the various models into three classes:

(a) If the model possesses a global chiral U_1 symmetry (with an anomaly) then $\mathscr{L}_\theta$ can be rotated away (i.e., $c = 0$) [2.139]. However, this possibility is disfavored, because if the U_1 is spontaneously broken there will be an axion [2.138] (a pseudo-Goldstone boson which acquires a very small mass from instanton effects). Axions are disfavored experimentally [2.141], although there are ways to evade the experimental limits [2.142]. Alternately, if the U_1 is not spontaneously broken then there must be a massless quark. However, the current quark masses ratios

$$\frac{m_u}{m_d} = 0.47 \pm 0.11$$
$$\frac{m_d}{m_s} = 0.042 \pm 0.007 \qquad (2.229)$$

obtained [2.143] from the meson and baryon mass spectra, ρ-ω mixing, and the $\eta \to 3\pi$ decay rate are incompatible with $m_u = 0$ or $m_d = 0$. (The small values of m_u and m_d tend to suppress d_N, but this is included in the estimate of c [2.144].)

(b) In models without a U_1 symmetry, if the CP violation is hard (i.e., if there are dimension four terms in the Lagrangian which violate CP, such as $\mathscr{L}_\theta$, complex Yukawa terms, or complex quartic Higgs couplings) then both θ_{QCD}

and θ_{QFD} receive divergent renormalizations. One can simply choose the renormalized θ to be sufficiently small, but this is not very elegant, especially since it typically requires that θ_{QCD} be chosen to cancel θ_{QFD}, and it is not obvious why these two parameters should be related. (The situation is similar to the problem with the cosmological constant, although not nearly so severe.) Ellis and Gaillard have argued [2.145] that θ is in fact a momentum dependent quantity. In the KM version of the GWS model, which has hard CP violation, the divergent renormalizations do not occur until very high order. Therefore, the Q^2 dependence of θ is very slow [2.145], and if θ is very small or zero for some unknown reason for any mass scale between zero and the Planck mass it will remain small in the entire range. Analogous statements do not hold in most other models of hard CP violation.

(c) Another possibility is to impose CP as a symmetry of the Lagrangian, or at least of the dimension four terms, so that CP is violated spontaneously (in the vacuum) or softly (through terms of dimension <4). In this case the bare value of θ_{QCD} is zero. The corrections and θ_{QFD} are finite and calculable.

Spontaneous CP violation means that the Lagrangian is CP invariant but some of the Higgs fields have complex VEV's. This can occur if there are two or more Higgs multiplets [2.146] (three are required if a symmetry $\varphi_m \rightarrow -\varphi_m$ is imposed [2.130]). These VEV's generate CP violating phases in the gauge and Yukawa couplings when they are expressed in terms of the mass eigenstate fermion and Higgs fields. From the magnitude of CP violation observed in the kaon system, one typically expects $\theta_{QFD} \simeq 10^{-3}$, which is much too large. Several authors [2.132-133,2.147] have constructed models in which extra symmetries (in addition to CP), which may be discrete, or continuous (global or local), are imposed on the Lagrangian in order to restrict the form of the Yukawa couplings and force det m_L to be real at the tree level. Of course, a non-zero $\bar{\theta}_{QFD}$ is still generated by loop corrections to the quark mass matrix, but in the examples cited it can be sufficiently small. In some cases the extra symmetries force the CP violating phase in the KM matrix to vanish, so that physical CP violation must be mediated by Higgs or additional gauge bosons (charged or neutral).

None of the solutions to the strong CP problem appear completely satisfactory. Models with hard CP violation require an artificial adjustment of the renormalized θ. Models with spontaneous or soft CP violation generally require a complicated Higgs structure and apparently ad hoc extra symmetries. As will be further discussed in Chapters 5 and 6, these extra symmetries are often incompatible with the constraints of grand unification. Furthermore, in most models with spontaneous CP violation the CP invariance is restored at high temperature so that a baryon asymmetry cannot be generated (some exceptions are discussed in Chapter 6). The restoration of CP at high temperature could also lead to an undesirable domain structure of the universe [2.148].

Conclusion

In conclusion, there are an enormous variety of extensions of the GWS model, including the possible existence of new gauge bosons, new fermions, and additional Higgs bosons. Many grand unified theories incorporate some or all of these extensions, although they are also possible outside of the context of grand unification. Tests for such effects include increased precision in the measurement of ordinary charged and neutral current processes, improved studies of CP violation, studies of the production and decay of heavy fermions, and searches for neutrino oscillations. Especially important are improved searches for flavor changing neutral current effects, such as in rare decays of kaons and leptons.

2.5 The Strong Interactions

2.5.1 Description

The basic properties of the strong interactions, as determined by hadron-hadron interactions, are: (a) They are very strong. For example, the pion nucleon coupling constant $g_{\pi NN}$ is given by $g^2_{\pi NN}/4\pi \simeq 14$. (b) They are short ranged ($r \simeq 1$ fm). (c) Hadrons are big, with sizes of order 1 fm. (d) They are invariant w.r.t. C, P, and T transformations, and they conserve I^3, strangeness, charm, electric charge, baryon number, etc. (e) They are approximately invariant w.r.t. global SU_2, SU_3, chiral SU_2, and chiral SU_3 transformations.

Many phenomenological models and theoretical ideas approximately describe and correlate various aspects of the strong interactions. For example, one boson exchange potentials, when supplemented with hard or soft cores, give a reasonable description of the low energy NN interaction (Fig. 2.10). High energy scattering is described by Regge theory, dual models, and other S-matrix type ideas. Soft pion theorems and current algebra sum rules, which are closely related to chiral symmetry, constrain low energy meson-nucleon interactions and various hadronic weak decays. The spectrum and properties of hadronic resonances are partially described by SU_3 symmetry and Regge theory.

The quark model [2.149] describes many more aspects of the spectrum, decays, and interactions of the hadrons. Persuasive additional evidence for the quark picture was provided by the scaling behavior observed in the deep inelastic reaction $ep \to e + X$ (Fig. 2.11). Scaling means that the cross section remains large as the momentum transfer $Q^2 \equiv -q^2 \to \infty$, as is expected if the nucleon consists of non-interacting point-like constituents (quarks, or as they were originally called in this context, partons), rather than falling rapidly as would be expected if the nucleon were large and diffuse. The details of the scattering supported the spin-$\frac{1}{2}$ nature of the constituents [2.150].

There are two distinct distance scales relevant to the strong interactions. At short distances (high momentum transfers) hadrons appear to consist of weakly or non-interacting quarks. At long distances (low momentum transfers), on the other hand, the hadrons appear to be large and strongly interacting. Also, isolated quarks have never been definitely observed [2.151] (although there are some indications in the Fairbank experiment [2.152]). These facts suggest that the strength of the interaction between quarks is zero or very small when the quarks are close together (asymptotic freedom) but gets large when the quarks are far apart. It would therefore be impossible or very difficult for quarks to be isolated from each other (confinement).

Each flavor of quark is believed to exist in three varieties or colors [2.153]. Physical hadrons are neutral w.r.t. the color quantum number. The color quantum number was first motivated by the statistics problem in the baryon spectrum (Section 2.2). Additional evidence includes the $\pi^0 \to 2\gamma$ decay rate and the e^+e^- annihilation cross section.

All of these aspects of the strong interactions (symmetry properties, quarks, colors, asymptotic freedom, etc.) are tied together in quantum chromodynamics (QCD), which is the only serious candidate for a field theory of the strong interactions.

2.5.2 Quantum Chromodynamics [2.154]

Quantum chromodynamics [2.155] (QCD) is a gauge theory based on the gauge group SU_3. SU_3 has eight generators, and the structure constants $c_{ijk} = f_{ijk}$, $i,j,k = 1,\ldots,8$ [2.156] are totally antisymmetric. SU_3 acts on the quark color indices (without changing flavor), so it is often written as SU_3^c, where c represents color. The quark fields q^α of a definite flavor transform according to the fundamental triplet representation $L^i = \lambda^i/2$, $i = 1,\ldots,8$, where the λ^i are the Gell-Mann matrices [2.156]. The leptons are of course SU_3^c singlets. SU_3^c is assumed to not be spontaneously broken. Therefore the eight gauge fields G^i (referred to as gluons) are massless. It is often convenient to express the gluon fields in a non-Hermitian basis defined by $G^\alpha_\beta \equiv (G)_{\alpha\beta}$, where the 3×3 matrix G is defined by

$$G \equiv \sum_{i=1}^{8} G^i \frac{\lambda^i}{\sqrt{2}} , \tag{2.230}$$

so that $(G^\alpha_\beta)^\dagger = G^\beta_\alpha$ and $G^\alpha_\alpha = 0$.

The QCD Lagrangian is

$$\mathscr{L}_{QCD} = -\frac{1}{4} F^i_{\mu\nu} F^{\mu\nu i} + \sum_r \bar{q}^r_\alpha \, i \, \not{D}^\alpha_\beta \, q^\beta_r - (\bar{q}_{L\alpha} \, m^o \, q^\alpha_R + \text{H.C.}) , \tag{2.231}$$

plus ghost terms. (One can also add $\mathscr{L}_\theta$ defined in (2.226).) The field tensor is given by

$$F^i_{\mu\nu} = \partial_\mu G^i_\nu - \partial_\nu G^i_\mu + g_s f_{ijk} G^j_\mu G^k_\nu . \tag{2.232}$$

The gluon kinetic energy term is rewritten in the G^α_β basis in Section 3.2. In (2.131), the index r runs over the quark flavors. The quark covariant derivative

$$
\begin{aligned}
D^{\alpha}_{\mu\beta} &\equiv (D_\mu)_{\alpha\beta} \\
&\equiv \partial_\mu\, \delta_{\alpha\beta} - i\, g_s\, G^i_\mu\, (L^i)_{\alpha\beta} \qquad (2.233) \\
&= \partial_\mu\, \delta_{\alpha\beta} - \frac{i\, g_s}{\sqrt{2}}\, G^{\alpha}_{\mu\beta}
\end{aligned}
$$

is independent of flavor and non-chiral (the same for q_L and q_R). m^o, the bare (or current) quark mass matrix, is independent of color but is a matrix in flavor space. m^o may actually be generated by the Higgs mechanism in the weak sector of the theory, but it can be thought of as a bare term as far as QCD is concerned. The form of the quark covariant kinetic energy term is unchanged by the A_L and A_R transformations that diagonalize m^o, so without loss of generality we may take m^o to be real and diagonal. The diagonal entries m^o_r are the bare or current masses of quark flavor r.

Asymptotic Freedom

The SU^c_3 gauge coupling constant g_s and the strong fine structure constant $\alpha_s \equiv g_s^2/4\pi$ determine the strength of the interaction between two quarks, as indicated in Fig. 2.12. Actually, α_s is not a constant, but a function of $Q^2 \equiv -q^2$, where q_μ is a typical momentum relevant to the process being considered. To see this intuitively, consider the higher order vacuum polarization diagrams to the gluon propagator shown in Fig. 2.12b-c. The virtual quark-antiquark pair in 2.12b will screen the color force, while the virtual gluons in 2.12c will anti-screen. The effective color interaction strength will therefore be a function of the distance between the quarks or, equivalently, of the momentum carried by the gluon. If the number of quark flavors is not too large, the antiscreening effects will dominate and the interaction will become weaker for high momentum (short distances). Therefore, QCD incorporates asymptotic freedom [2.157]. Non-abelian gauge theories are the only realistic asymptotically free field theories [2.157-158].

A proper calculation of the effective or running coupling $\alpha_s(Q^2)$, including vertex and quark self energy diagrams, as well as a precise definition of its meaning, is given in [2.157]. The equation satisfied by α_s is

$$\frac{d\alpha_s}{d\,\ell n\, Q^2} = 4\pi\, b\, \alpha_s^2 + O(\alpha_s^3) \;, \tag{2.234}$$

where

$$b = -\frac{1}{(4\pi)^2}\left[11 - \frac{2n_q}{3}\right] \;, \tag{2.235}$$

where $n_q = 2F$ is the number of quark flavors (F is the number of families). Actually, only those quarks which are light compared to $\sqrt{Q^2}$ are counted in n_q. For $n_q \leq 16$, we have $b < 0$, which means that the theory is asymptotically free. The solution of (2.234) is (neglecting the α_s^3 terms)

$$\frac{1}{\alpha_s(Q^2)} = \frac{1}{\alpha_s(\Lambda^2)} - 4\pi\, b\, \ell n\, \frac{Q^2}{\Lambda^2} \tag{2.236}$$

where Λ is an arbitrary reference momentum. For $Q^2 \to \infty$, the first term on the right can be ignored, giving

$$\alpha_s(Q^2) = \frac{12\pi}{33-2n_q}\,\frac{1}{\ell n\,\frac{Q^2}{\Lambda^2}} \;, \qquad \xrightarrow[Q^2\to\infty]{} 0 \tag{2.237}$$

which displays the asymptotic freedom of QCD for $n_q \leq 16$.

Therefore, QCD incorporates asymptotic freedom and approximate scaling (the parton model) for $Q^2 \to \infty$. For small Q^2 (long distances), on the other hand, the right-hand side of (2.237) becomes large. The perturbation theory approximations used in deriving (2.237) break down in this regime, and one can only speculate on what happens. One possibility is that the true α_s grows large for small Q^2, resulting in the strong coupling regime of strong interactions, quark confinement, and more generally the confinement of all fields that carry color. If this last hypothesis is correct, then gluons cannot propagate freely through space. This would be one reason that the observed strong interactions are short ranged. Also, physical hadrons are color singlets. Thus, they cannot emit or absorb a single gluon, but must interact via the analog of dipole-dipole forces. Two nucleons could interact by the exchange of two or more gluons or of a $q\bar{q}$ pair, for example. The one boson exchange model of the N-N interaction would then be an approximation in which the $q\bar{q}$ pairs are assumed to form bound state mesons.

Let us temporarily consider QCD without the heavy quarks (c,b,...), which presumably have little relevance to ordinary hadrons. If we further neglect the bare masses of the light quarks (u,d,s) (this should be a good first approximation, at least for the u and d quarks) then $\mathscr{L}_{QCD}$ contains no dimensional parameters. In fact, the renormalized effective fine structure constant $\alpha_s(Q^2)$ in (2.237) depends only on an arbitrary reference mass Λ, not on the bare gauge coupling in $\mathscr{L}_{QCD}$. Therefore, in the limit of neglecting bare masses QCD has no arbitrary parameters (other than the CP violating θ parameter discussed in Section 2.4.4). α_s depends only on Q^2/Λ^2; the other hadronic mass scales, such as the proton mass and the pion decay constant, are presumably given by pure numbers times Λ (of course, we don't know how to calculate the coefficients). In some sense, the Lagrangian coupling constant g_s has been traded for a massive parameter Λ. The value of Λ is not observable, however, since it merely sets the mass scale for the strong interactions. This picture is not greatly altered by the introduction of nonzero masses for the u, d, and s quarks as long as they are small compared to Λ.

Since we do not know how to calculate the proton mass m_p in terms of Λ, it is necessary to invert the logic and use m_p or some related scale such as 1 GeV, as our reference scale. Then Λ in GeV can be found from $\alpha_s(Q^2)$, which in turn is determined from the deviations from scaling observed in deep inelastic reactions or from charmonium. There is a complication, however, in that there is no unique way to define the renormalized coupling $\alpha_s(Q^2)$. It cannot be defined in terms of the on mass shell quark-gluon vertex, for example, because quarks and gluons are presumably confined and cannot be "on mass shell." Various definitions of α_s, which differ in the finite parts of the renormalizations involved [2.159-161], correspond to different values of Λ. The expressions for physical quantities are independent of the renormalization prescription to lowest nontrivial order in α_s, but not to higher order. A convenient definition of α_s for deep inelastic scattering, which reduces the effects of higher order terms, is the modified minimal subtraction ($\overline{MS}$) scheme, in which not only the $1/\varepsilon$ poles in dimensional regularization but the associated factors of $\ell n\ 4\pi - \gamma_E$ are subtracted from the unrenormalized gauge coupling [2.160]. Values of Λ from various deep inelastic scaling violation measurements have been compiled by Ellis [2.162] and expressed in terms of $\Lambda_{\overline{MS}}$. He concludes that $\Lambda_{\overline{MS}} = 0.5$ GeV with an unknown (perhaps of order several hundred MeV) error. Higher twist effects could reduce $\Lambda_{\overline{MS}}$ by as much as a third.

Λ has also been determined from various aspects of the charmonium and bottomonium systems. Typical values obtained are [2.163] $\Lambda \lesssim 100$ MeV. The apparent discrepancy with the value obtained from deep inelastic scattering may be due to the fact that higher order corrections have not been included in the quarkonia estimates. Until such corrections are computed one cannot tell whether Λ should be identified with $\Lambda_{\overline{MS}}$ or with some other definition of Λ. The experimental and theoretical status of the determination of Λ leaves much to be desired. This is unfortunate, because Λ is a principal input into the estimate of the proton lifetime in grand unified theories (Chapter 4).

Running Coupling Constants

It will be useful for future reference to list the relevant formulas for the effective or running coupling constants $g(Q^2)$ and $\alpha = g^2/4\pi$ in a general gauge theory. Only the effects of gauge bosons and fermions (both of which must be light compared to the momentum to be included in the formulas) are given. For the effects of scalars, see [2.157-158,2.164]. The key equation is

$$\frac{d\, g^2}{d\,\ell n\, Q^2} = b\, g^4 + O(g^6) \ , \tag{2.238}$$

where

$$b = \frac{-1}{(4\pi)^2}\left[\frac{11}{3}\, C_2(G) - \frac{4}{3}\, T_f\right] \ ; \tag{2.239}$$

$C_2(G)$, the quadratic Casimir operator for the adjoint representation of G, is given by

$$C_2(G)\ \delta_{ij} = \sum_{k,\ell} c_{ik\ell}\, c_{jk\ell} \ . \tag{2.240}$$

For example, $C_2(SU_n) = n$ and $C_2(U_1) = 0$. T_f, which is related to the Casimir operator for the fermion representation, is

$$T_f \, \delta_{ij} = \frac{1}{2} \mathrm{Tr}(L_L^i \, L_L^j) + \frac{1}{2} \mathrm{Tr}(L_R^i \, L_R^j) \xrightarrow[L_L = L_R = L]{} \mathrm{Tr} \, L^i \, L^j \,, \tag{2.241}$$

where $L_{L,R}$ are the representation matrices for the left- and right-handed fermions, respectively. If b is positive (negative) g increases (decreases) as Q^2 increases. The solution of (2.238) is (ignoring the g^6 terms)

$$\frac{1}{\alpha(Q^2)} = \frac{1}{\alpha(M^2)} + 4\pi \, b \, \ln \frac{M^2}{Q^2} \,, \tag{2.242}$$

where M is a reference scale, such as the Λ used in QCD. However, for applications to GUTS it will be convenient to take M equal to the grand unification scale.

For SU_3^c, we have $C_2(G) = 3$ and $T_f = n_q/2$, where $n_q = 2F$ is the number of color triplets (flavors). For SU_2 (in the GWS model), $C_2 = 2$ and $T_f = n/4$, where n is the number of left-handed doublets. Hence, $n = 4F$, since each family has one lepton doublet and quark doublets for each of the three colors. For U_1^Y, $C_2(G) = 0$ and $T_f = \frac{1}{2} \mathrm{Tr} \, Y_L^2 + \frac{1}{2} \mathrm{Tr} \, Y_R^2 = \frac{5F}{3}$. Therefore, for $Q^2 \gg M_W^2$, M_Z^2, we have

$$\begin{aligned}
\frac{d\,\alpha_s^{-1}}{d \ln Q^2} &= \frac{1}{4\pi} \left[11 - \frac{4}{3} F \right] \\
\frac{d\,\alpha_g^{-1}}{d \ln Q^2} &= \frac{1}{4\pi} \left[\frac{22}{3} - \frac{4F}{3} \right] \\
\frac{d\,\alpha_{g'}^{-1}}{d \ln Q^2} &= \frac{1}{4\pi} \left[- \frac{20F}{9} \right] .
\end{aligned} \tag{2.243}$$

The SU_3^c and SU_2 gauge couplings are asymptotically free for $F < 9$ and $F < 6$ respectively. $\alpha_{g'}$ increases with Q^2 for any $F > 0$. Similarly, for $Q^2 < M_W^2$, M_Z^2 the unbroken U_1 group is U_1^{EM}, for which $T_f = \mathrm{Tr}\, Q^2 = 8F/3$ (ignoring fermion masses), so that

$$\frac{d\,\alpha_e^{-1}}{d\,\ln Q^2} = \frac{1}{4\pi}\left[-\frac{32F}{9}\right] < 0 \qquad (2.244)$$

Symmetries of QCD

Most of the symmetries and conservation laws respected by the strong interactions are automatic consequences of QCD. That is, the constraint of renormalizability (gauge invariance) completely determines the form of $\mathscr{L}_{QCD}$, except for the number of quark flavors, the mass matrix m^o, and the P and CP violating $\mathscr{L}_\theta$ (2.226). The effects of $\mathscr{L}_\theta$ vanish to all finite orders in perturbation theory, but will generally be non-zero when non-perturbative effects are considered. I will assume that $\mathscr{L}_\theta$ is small or negligible for some unknown reason, as discussed in Section 2.4.4.

Then, $\mathscr{L}_{QCD}$ automatically conserves quark number N_q (baryon number is $N_q/3$). Furthermore, the only term in $\mathscr{L}_{QCD}$ that can violate C, P, T, or the quark number for individual flavors is the mass matrix m^o. However, m^o can be transformed into a real and diagonal form without affecting the rest of $\mathscr{L}_{QCD}$ (except for $\mathscr{L}_\theta$) so that without loss of generality we can work in the mass eigenstate basis. Then, C, P, T, I^3, Q, strangeness, charm, bottom number, top number, etc., are all automatically conserved by QCD. One can therefore regard the many symmetries of the strong interactions as dynamical accidents: no observable symmetry violation terms can be added to $\mathscr{L}_{QCD}$ without destroying renormalizability. One might fear that weak interaction corrections to strong processes might induce large parity violating effects. That is, a virtual $W^{\pm}$ boson in a loop diagram may induce parity violation in strong processes to order g^2 rather than to order G_F. Weinberg has shown [2.165], however, that for gauge theories such as QCD these parity violating $O(g^2)$ corrections only affect m^o. They can therefore be rotated away and are unobservable in strong processes.

QCD also incorporates current algebra and chiral symmetry ideas. The following scenario is generally believed, although many aspects have not been proven. In the limit of neglecting m^o, $\mathscr{L}_{QCD}$ possesses an exact global $SU_{n_q} \times SU_{n_q} \times U_1^B \times U_1^A$ chiral symmetry, where n_q is the number of quark flavors. U_1^B is just the baryon number symmetry, while U_1^A is the axial baryon number symmetry generated by the current

$$J_\mu^A = \sum_r \bar{q}_r \gamma_\mu \gamma_5 q^r \qquad (2.245)$$

The observed strong interactions do not exhibit the U_1^A symmetry. (This is the U_1 problem.) It is generally believed [2.166] that the U_1^A is violated by non-perturbative instanton effects which generate non-zero matrix elements for the anomaly in $\partial \cdot J^A$, but the issue is still controversial [2.167]. The chiral $SU_{n_q} \times SU_{n_q}$ group is generated by $n_q^2 - 1$ vector and $n_q^2 - 1$ axial vector charges which rotate the left- and right-handed flavors separately. The bare masses of the c and heavier quarks are believed to be sufficiently large that only the $SU_3 \times SU_3$ subgroup associated with the u, d, and s quarks is a good approximate symmetry of $\mathscr{L}_{QCD}$.

$SU_3 \times SU_3$ is explicitly broken by m_u^o, m_d^o, and m_s^o, which are called the bare or current masses. In the limit $m_u^o = m_d^o = m_s^o = 0$, the physical hadronic states are arranged in degenerate multiplets of the SU_3 generated by the vector charges. However, the physical spectrum does not exhibit any indication of parity doubling or massless fermions. The eight axial generators must therefore be spontaneously broken, implying the existence of eight massless pseudoscalar Goldstone bosons, which can be identified with the π, K, and η mesons (if U_1^A were a good symmetry there would be a ninth Goldstone boson with mass comparable to the pion after the quark masses are given non-zero values). There are several manifestations of this spontaneous symmetry breaking, including a non-zero VEV for the scalar operator $\bar{q}_r\, q_r$ (the analog of a Higgs field), a non-zero pion decay constant f_π, defined by the matrix element of the axial current between the vacuum and the one pion (or one kaon or η) state, and a non-zero constituent quark mass M^c (which is the same for u, d, and s). The constituent mass is a non-perturbative dynamical effect generated by the SSB. It would be the physical quark mass if quarks were not confined. With confinement, M^c is related to the mass parameters appearing in potential and bag models. It sets the scale for the nucleon masses. f_π and M^c, which measure the amount of spontaneous symmetry breaking, are of the order of hundreds of MeV. The scale is presumably set by Λ.

When the current masses are given non-zero but equal values, the chiral symmetry is explicitly broken down to ordinary SU_3. The major effect is that the Goldstone bosons acquire masses $\mu^2 \propto m$. For $m_u^o = m_d^o \neq m_s^o$, SU_3 is broken, and for $m_u^o \neq m_d^o$, SU_2 is explicitly broken, leading to splittings in the hadron multiplets (also the constituent masses are shifted). The absolute values of the current masses depend on the renormalization prescription, but the ratios are essentially independent of renormalization effects [2.143]. From the pseudoscalar and baryon spectra, ρ-ω mixing, and the $\eta \to 3\pi$ decay one obtains [2.143] $m_u/m_d = 0.47 \pm 0.11$ and $m_d/m_s = 0.042 \pm 0.007$ (m without the superscript refers to the renormalized current quark mass). Typical estimates [2.168] of the scale yield m_s in the range from 150 to 300 MeV, corresponding to m_u and m_d in the ranges 3-6 and 6-13 MeV, respectively. m_u and m_d are very small compared to other hadronic mass scales, so that $SU_2 \times SU_2$ is an excellent approximate symmetry of the strong interactions [2.169]. Also $m_u \neq m_d$ so that isospin symmetry is broken not only by electromagnetic effects but by quark mass differences in $\mathscr{L}_{QCD}$. The approximate validity of isospin is therefore not due to a near degeneracy of m_u and m_d but rather to the fact that they are both too small to greatly affect hadronic physics. The explicit breaking of SU_3 and of $SU_3 \times SU_3$ chiral symmetry by bare quark mass terms in QCD are concrete realizations of the Coleman-Glashow tadpole model [2.170] and of the Gell-Mann-Oakes-Renner model of chiral symmetry breaking [2.171].

Like coupling constants, the effective quark masses vary with Q^2. The current and dynamical masses are believed [2.171a] to go like a constant and $1/Q^2$, respectively, for large Q^2, up to logarithms.

Conclusion

QCD is an attractive theory of the strong interactions, which incorporates the quark model, asymptotic freedom, chiral symmetry, the one boson exchange model, and other desirable features, at least qualitatively. Many aspects, especially in the strong coupling regime, are beyond our ability to calculate in detail, but the qualitative features are encouraging. The only quantitative tests of QCD to date involve scaling violation in short distance processes. Even here the situation is obscured by higher order and higher twist corrections.

The basic ingredients of QCD are quarks, gluons, and their interactions. Few physicists today doubt the existence of quarks. The distribution of hadrons observed in e^+e^- annihilations at PETRA strongly suggests the existence of gluons. The observed strong interactions provide evidence for the quark-gluon interaction, but there is no direct evidence for the gluon self interaction, which is a signature of the non-abelian nature of the theory.

Other aspects of QCD, such as glueballs (which would provide evidence for the gluon self interaction), multi-quark states, jets, strings, bags, instantons, etc., are thoroughly discussed in the current literature.

I will briefly mention several possible modifications of QCD which are relevant to grand unification: (a) Many authors [2.172] have discussed the possibility of exotic heavy quark states placed in color sextets, octets, or other SU_3^c representations. Pati and Salam have emphasized the possibilities (b) that quarks may have integer electric charges (±1,0) [2.173] (in which case Q does not commute with SU_3^c; and (c) that the underlying color group is SU_4 [2.174], which is spontaneously broken down to SU_3^c. The leptons are identified as the fourth color, with (u^R,u^G,u^B,ν_e), (d^R,d^G,d^B,e^-), etc. in SU_4 quartets. One can consider both integer and fractional quark charge versions of this model. These models are considered in more detail in Section 3.4.5. (d) Okubo [2.175] has discussed other possible color groups.

2.5.3 Dynamical Symmetry Breaking

Motivations

Dynamical symmetry breaking (DSB) refers to the possibility of the spontaneous breaking of global or local symmetries without the introduction of explicit Higgs fields. An example is the presumed DSB of chiral symmetry in QCD.

There are several reasons that many people object to the introduction of Higgs fields into the standard model:

(a) They introduce many arbitrary free parameters, in the Higgs potential and Yukawa couplings, into the theory. This tends to undo one of the most attractive features of gauge theories, namely that they prescribe the form of the interactions.

(b) Higgs and Yukawa effective couplings tend to increase with Q^2 [2.157-158,2.164] (i.e., they tend to destroy asymptotic freedom).

(c) Scalar self energies are quadratically divergent, so that renormalized and unrenormalized scalar masses are related by

$$\mu^2 = \mu_o^2 + \lambda \kappa^2 , \tag{2.246}$$

where λ is the scalar quartic coupling and κ is an ultraviolet cutoff. Wilson and Susskind have argued [2.176] that κ should be interpreted as the Planck mass 10^{19} GeV (i.e., that quantum gravity somehow cuts off the divergence). The bare parameter μ_o^2 must then be adjusted (or fine tuned) to $\simeq$38 decimal places in order to obtain a small renormalized mass such as $\mu \simeq 1$ GeV! (Weinberg [2.177] has argued against this interpretation, however, on the grounds that (2.246) is simply an artifact of the regularization procedure.)

(d) In most grand unified theories there are two distinct mass scales, at $M_W \simeq 100$ GeV and at 10^{14} GeV. This large ratio or hierarchy of masses is difficult to obtain using the Higgs mechanism.

Technicolor

The basic idea in most DSB schemes is to replace the Higgs field by a composite field [2.178-180]. Let us first consider the example given by

Susskind [2.180] of an ordinary $SU_3^c \times SU_2 \times U_1$ model without Higgs fields. Presumably, the composite operators $\bar{u}u$ and $\bar{d}d$ develop non-zero equal vacuum expectation values

$$<0|\bar{u}u|0> = <0|\bar{d}d|0> \simeq \Lambda^3 \tag{2.247}$$

where $\Lambda \simeq M^c$ is the QCD parameter which describes the scale at which the strong interactions become strong. These non-zero VEV's, or vacuum condensates, imply the DSB of $SU_2 \times SU_2$ chiral symmetry. There will be three composite Goldstone bosons π^i, $i = 1,2,3$. The vacuum condensate also breaks the $SU_2 \times U_1$ electroweak gauge symmetry. The vacuum polarization tensors of the W and Z bosons will develop poles at $k^2 = 0$ due to the Goldstone pions as shown in Fig. 2.13, with residues $g^2 f_\pi^2/4$ and $(g^2 + g'^2)f_\pi^2/4$, respectively, where the pion decay constant (which is related to M^c) is defined by

$$<0|\bar{q}\, \gamma_\mu\, \gamma_5\, \frac{\tau^i}{2}\, q|\pi^j(k)> = i\, k_\mu\, f_\pi\, \delta^{ij} . \tag{2.248}$$

Therefore, the W and Z bosons will acquire masses [2.180]

$$M_W^2 = M_Z^2 \cos^2\theta_W = \frac{g^2}{4} f_\pi^2 . \tag{2.249}$$

The pion degrees of freedom are "eaten" by the gauge bosons, so that the composite pions play the role of the Higgs fields, with the pion decay constant, which measures the amount of DSB, replacing the VEV ν in (2.158). The relation between M_W and M_Z is preserved by the DSB because of the unbroken SU_2 symmetry of the condensate, which ensures equal decay constants for all three pions.

This model cannot be considered a realistic model of DSB because it leads to $M_W \simeq 30$ MeV (i.e., $f_\pi/\nu \simeq 1/3000$). Also, there are physical pions in the real world, so they have not been eaten.

Weinberg [2.179] and Susskind [2.180] have suggested the existence of a new interaction, called technicolor (TC) or hypercolor [2.181], or primed color [2.122], which is assumed to be similar to the color interaction except that its scale parameter $\Lambda' \simeq 1$ TeV is $\approx 3000\times$ larger than the scale Λ of QCD. (It need not be based on the SU_3 group, however.) Ordinary quarks and leptons would be singlets w.r.t. technicolor, while new families of fermions would be introduced to carry the TC quantum number. Some of these technifermions would transform as SU_2 doublets, and they could carry ordinary color as well. Suppose one introduces an SU_2 doublet of techniquarks, U and D, for example. Then, in analogy with QCD, it is assumed that the $SU_{2L} \times SU_{2R}$ symmetry of the techniquarks is dynamically broken by a condensate

$$<0|\overline{U}U|0> = <0|\overline{D}D|0> \simeq \Lambda'^3 \ . \tag{2.250}$$

The associated technipions π', with decay constants $F_\pi \simeq 3000\ f_\pi$, will then be eaten by the $W^\pm$ and Z bosons, yielding masses

$$M_W^2 = M_Z^2 \cos^2\theta_W = \frac{g^2}{4} F_\pi^2 \tag{2.251}$$

of the correct values. The techniquarks will have constituent masses of order $\Lambda' \simeq 1$ TeV. This should be the typical scale of bound states of the techniquarks (techni-mesons and -baryons).

The ordinary quarks and leptons, which are TC singlets, are not much affected by the TC interactions. In fact, the major problem of simple TC is that there is no mechanism for generating current algebra masses for the ordinary fermions. There should still be a condensate of the u and d quarks, generating three massless pions, which exist as real particles.

The dynamics of the TC scenario have not been established, although some progress has been made by Pagels [2.182]. An interesting alternative by

Marciano [2.183] dispenses with TC and introduces quarks transforming according to sextets or octets w.r.t. ordinary SU_3^c. He argues that these quarks may form condensates at mass scales much larger than those relevant to color triplets because of their larger color couplings.

Extended Technicolor (ETC)

Dimopoulos and Susskind [2.184] and Eichten and Lane [2.181] have suggested that the problem of generating current algebra masses for the ordinary quarks and leptons could be solved by embedding the TC group in a larger gauge group, the extended technicolor group (ETC), in which the new generators connect ordinary and TC fermions. For example, a quark q and a techniquark Q could be combined in a multiplet (q Q) of ETC, such that the TC subgroup acts only on Q while the new bosons, called the E bosons, of the ETC/TC quotient group connect the q and Q. One must assume that the ETC group is somehow spontaneously broken down to an unbroken TC subgroup, so that the E bosons obtain very large masses. Then quark q can acquire a mass from the diagram in Fig. 2.14, giving

$$
\begin{aligned}
m_q &\simeq \frac{g_E^2}{(2\pi)^4} \frac{1}{M_E^2} \int_o^{m_Q^2} \frac{d^4\ell \; m_Q}{\ell^2 + m_Q^2} \\
&\simeq \frac{g_E^2}{8\pi^2} \frac{m_Q^3}{M_E^2} \simeq \frac{g_E^2}{8\pi^2} \frac{\langle 0|\overline{Q}Q|0\rangle}{M_E^2} \quad ,
\end{aligned}
\tag{2.252}
$$

where the cutoff on the integral occurs because the running constituent mass of Q decreases rapidly for $\ell^2 \gtrsim m_Q^2$. One expects $g_E^2/8\pi^2 \simeq 1$ at the relevant momentum scales so that for $m_q \simeq 100$ MeV and $m_Q \simeq 1$ TeV one must have $M_E \simeq 100$ TeV. m_q will be a constant (up to anomalous dimension logarithms neglected in (2.252)) for momenta small compared to m_Q, but will fall rapidly for $\ell^2 \gtrsim m_Q^2$. This should be sufficient for m_q to successfully imitate a true current algebra mass in ordinary low energy applications.

A number of comments are in order:

(a) The possibility of TC interactions and DSB is logically independent of the existence of elementary scalar fields. The ordinary GWS Higgs particles could exist in addition to new interactions, or TC could replace the GWS Higgs but not the Higgs fields associated with the breakdown of grand unified theories.

(b) One of the problems with ETC is that the ETC must be spontaneously broken at a scale of 100 TeV. This could be done by an explicit Higgs mechanism or by yet another level of superstrong interactions that give mass to the E boson in the same way that TC gave mass to the W and Z. These new interactions could be an extension of the ETC group [2.181,2.184-185] or somehow be the ETC interactions themselves [2.181].

(c) Raby, Dimopoulos, and Susskind [2.185] have suggested that a large ETC group may naturally break down in a hierarchy of steps, with typical masses M_{E_1}, M_{E_2}, ... (See Chapter 6.) This could explain the hierarchy of light fermion masses: the mass hierarchy of the E_i would be carried over (and magnified by the quadratic dependence) to the fermion masses generated by the exchange of E_i.

(d) Models in which color and technicolor are unified at a high momentum scale may solve the strong CP problem [2.186]. There is a single θ parameter in $\mathscr{L}$ that can be set equal to zero by an appropriate phase choice for the fermions. The Lagrangian is then CP invariant, and physical CP violation must somehow be generated dynamically by the condensates. Whether a realistic physical CP violation can occur without producing an unacceptably large effective θ is unknown.

(e) The successful relation $M_W^2 = M_Z^2 \cos^2\theta_W$ requires that the TC condensate be invariant w.r.t. an SU_2 group under which the SU_{2L} gauge bosons transform as a triplet [2.179-180,2.187]. If this SU_2 is the isospin group then in order to have $m_u \neq m_d$ it is necessary for the ETC couplings to explicitly violate isospin [2.181,2.184,2.187], e.g., by assigning u_R and d_R to different ETC representations [2.181].

(f) In ETC models with more than a single doublet of techniquarks there will typically be many Goldstone bosons [2.181,2.184], only three of which are eaten. The others will acquire masses in the 10-25 GeV range [2.181] ($\simeq \sqrt{\alpha}\, F_\pi$) by radiative corrections associated with the $SU_3^c \times SU_2 \times U_1$ interactions (which break the techni-chiral symmetry). These bosons preferentially couple to the heaviest fermions, but unlike the Higgs particle in the simplest GWS model they usually couple as pseudoscalars. Their properties are discussed in [2.181,2.184,2.188-189].

(g) The ETC interactions may mediate [2.181] observable flavor changing neutral current interactions of the light fermions with strength $\simeq M_E^{-2} \simeq 10^{-5}\text{-}10^{-6}\, G_F$.

Dynamical symmetry breaking is an extremely attractive idea, but so far no really compelling examples have been given. The ETC models manage to eliminate the Higgs doublet, but only at the cost of introducing an enormously complicated structure of new interactions.

Attempts to combine the DSB ideas with grand unified models will be discussed in Chapter 6.

2.6 The Standard Model

2.6.1 Description

The standard model of the strong, weak, and electromagnetic interactions is just the combination of the GWS model of electroweak interactions with QCD. The gauge group is the direct product $G_s \equiv SU_3^c \times SU_2 \times U_1$, with couplings g_s, g, and g' for the three factors. Ignoring the generalized Cabibbo mixings, the fermion representations are

$$\begin{array}{llll} \text{1st Family:} & \begin{pmatrix} u^\alpha \\ d^\alpha \end{pmatrix}_L & \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L & u_R^\alpha,\ d_R^\alpha,\ e_R^- \\ \text{2nd Family:} & \begin{pmatrix} c^\alpha \\ s^\alpha \end{pmatrix}_L & \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L & c_R^\alpha,\ s_R^\alpha,\ \mu_R^- \\ \text{3rd Family:} & \begin{pmatrix} t^\alpha \\ b^\alpha \end{pmatrix}_L & \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L & t_R^\alpha,\ b_R^\alpha,\ \tau_R^- \ , \end{array} \tag{2.253}$$

where SU_2 doublets are arranged in a column and SU_3^c triplets are indicated by the superscript α. The particle content could just as well have been expressed entirely in a basis of left handed fields using $\psi_L^c \equiv C\,\bar{\psi}_R^T$. The first family would then be

$$\begin{pmatrix} u^\alpha \\ d^\alpha \end{pmatrix}_L \quad \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad u_{L\alpha}^c\ d_{L\alpha}^c\ e_L^+ \ , \tag{2.254}$$

where the charge conjugate quark fields transform as 3^* under SU_3^c. One can add additional families and/or right handed neutrinos ν_R (or ν_L^c) to (2.253) if desired.

2.6.2 Unanswered Questions

The standard model successfully describes or is at least consistent with all known facts of elementary particle physics. It is a mathematically consistent field theory in which all of the known interactions are basically gauge interactions. (There are also the Higgs self interactions and the much weaker Yukawa interactions.) The standard model is probably correct to some level of approximation, but few physicists believe that it is really the ultimate theory of elementary particles. The problem is not that the model has any inconsistencies or wrong predictions, but that it has too much arbitrariness.

Amongst the unexplained features and unanswered questions are:

(a) The pattern of groups and representations is complicated and arbitrary. Why should the gauge group be a direct product of three different factors? Why are the fermion representations such that SU_3^c is non-chiral (parity conserving) while $SU_2 \times U_1$ is chiral? What is the purpose of the second and third families, which are merely more massive repetitions of the first family? (This is the modern version of the old question, "Why does the muon exist?")

(b) The strong, weak, and electromagnetic fine structure constants (at $Q^2 \simeq 10\ \mathrm{GeV}^2$) are $\alpha_s \simeq 0.4$, $\alpha_g \simeq 0.03$, and $\alpha_e \simeq 0.007$. Why do these couplings, and $\sin^2\theta_W = \alpha_e/\alpha_g$, take the values that they do, and in particular, why are they so different?

(c) Electric charge is not quantized. That is, the electric charge operator is $Q = T^3 + Y$, where the hypercharge assignment can be made independently for each representation. The only group theoretic constraint is that the charge differs by one unit between the fields that are associated in a doublet. However, the charges of leptons, quarks, and Higgs scalars need not be related by simple factors like one or three. (If one adds the further constraint that anomalies in the $A^i A^i B$ and B^3 vertices should be absent, then one has $3y_{qL} + y_{\ell L} = 0$ (the hypercharges of the right-handed fields are fixed by requiring $q_{eL} = q_{eR}$, etc.), but this does not suffice to relate the quark and lepton charges. If one also imposes $q_\nu = 0$, then $y_{\ell L} = -\frac{1}{2}$ and $y_{qL} = +\frac{1}{6}$, which yield the correct fermion charges. The Higgs charges are still unconstrained, however.)

(d) The Higgs parameters and the Yukawa couplings are free parameters. This means that the complicated pattern of fermion masses, the fermion mixing angles and phases, and the W and Higgs masses are completely arbitrary. The neutrino mass can be forced to be zero by not introducing a ν_R singlet, but if it is introduced there is no reason for the neutrino mass to be small. Even given the hypercharge assignments, the standard model for three families has 19 observable free parameters [2.60] (26 if right-handed neutrinos are introduced). These are: 3 gauge couplings, two θ parameters for the SU_3^c and SU_2 subgroups (one can add the analog of $\mathscr{L}_\theta$ to SU_2, but its effects are much too small to be observable [2.190]). The parameters μ^2 and λ from the Higgs potential (which determine M_W and μ_η), 10 parameters from the quark mass matrix (6 masses, 3 mixing angles, and one CP violating phase), and 3 charged lepton masses (with ν_R one must add 3 neutrino masses, 3 lepton mixing angles, and 1 phase), for a total of 20 (or 27). One must then subtract one overall mass scale to obtain 19 (or 26).

(e) There are several quantities that are arbitrary in the standard model which appear to be unnaturally small. These include the ratio of the neutrino masses (if they are non-zero) to other fermion masses, the ratio of fermion masses to the W and Z masses (i.e., the ratio of the Yukawa couplings to the gauge coupling), the ratio of the strong interaction scale Λ to the weak interaction scale M_W, the value of the θ parameter, and the incredibly small ratio of the observed cosmological term to the value induced by SSB (the primordial cosmological term could have been added to the enumeration of free parameters above). The values of these quantities suggest the existence of dynamical constraints that force them to be small (some possibilities for θ were discussed in Section 2.4.4).

(f) Gravity has not been included in the standard model.

Grand unified theories, to which I now turn, were motivated in part by the desire to constrain some of these quantities that are arbitrary at the level of the standard model.

3. GRAND UNIFIED THEORIES

3.1 General Description

The basic idea in a grand unified theory is that if $G_s \equiv SU_3^c \times SU_2 \times U_1$ is embedded in a larger underlying group G then the additional symmetries may restrict some of the features that were arbitrary in the standard model. A typical consequence of this embedding is that the new symmetry generators and their associated gauge bosons involve both flavor and color. The new interactions generally violate the conservation of baryon number and, in most models, lead to proton decay. The observed limit of $\tau_p > 10^{30}$ yr on the proton lifetime then requires that the baryon number violating interactions must be extremely weak. For models in which the proton can decay via the exchange of a single gauge boson X, the lifetime limit typically requires that $M_X > 10^{14}$ GeV, twelve orders of magnitude larger than the $W^{\pm}$ and Z masses! An unfortunate consequence of this extreme weakness is that the experimentally accessible consequences of the new interactions, other than proton decay, are very few or nonexistent.

If G is simple, which basically means that it is not a direct product of factors like SU_3 or SU_2, or if it is a direct product of identical simple groups related by discrete symmetries, then G has only one gauge coupling constant. If the theory is probed at momenta Q^2 large compared to M_X^2, where all spontaneous symmetry breaking effects can be ignored, then the strong, weak, electromagnetic, and baryon number violating interactions all look basically similar and there is a single coupling constant. Quarks, anti-quarks, leptons, and anti-leptons would be fundamentally similar; typically, some or all of these are placed together in the same representation of G. It is only for $Q^2 < M_X^2$ that SSB becomes important and the running fine structure constants α_3, α_2, and α_1 of the SU_3, SU_2, and U_1 subgroups become different, as shown in Fig. 3.1. The observed strong, weak, and electromagnetic interactions, with their very different properties and coupling constants, are therefore simply the result of the pattern of SSB of the underlying group G.

The values of the coupling constants measured at $Q^2 \lesssim M_{W^\pm}^2$ can be used in the simpler models to predict M_X, the mass of the boson which mediates proton decay. Because of the large difference between the couplings at low momentum and their slow (logarithmic) Q^2 variation, M_X is typically predicted to be extremely large. In the Georgi-Glashow model, for example, M_X is predicted to be $\gtrsim 10^{14}$ GeV, the same scale needed to explain the approximate stability of the proton!

3.2 The Gauge Group SU_n

Before discussing the Georgi-Glashow SU_5 model, it will be useful to review the formalism of the SU_n groups. The group SU_n is defined by its fundamental representation, which is the group of $n \times n$ unitary matrices (with complex entries) with determinant one. A general SU_n transformation can be written as

$$\begin{aligned} U &= \exp\left(-i \sum_{i=1}^{n^2-1} \beta^i L^i\right) \\ &= \exp(-i\,\vec{\beta} \cdot \vec{L}) , \end{aligned} \tag{3.1}$$

where the n^2-1 generators L^i are Hermitian (which quarantees that U is unitary) and traceless (which implies det U = 1). The L^i may be normalized so that $\mathrm{Tr}(L^iL^j) = \delta_{ij}/2$. SU_n has rank n - 1, which means that n - 1 of the generators can be simultaneously diagonalized. It is conventional to write $L^i \equiv \lambda^i/2$; for the case $n = 2$, the λ^i are the Pauli matrices.

A convenient basis for the representation matrices can be defined in terms of the traceless matrices L^a_b defined by

$$(L^a_b)_{cd} \equiv (L^a_b)^c{}_d \equiv \delta^c_b\,\delta^a_d - \frac{1}{n}\delta^a_b\,\delta^c_d \tag{3.2}$$

(the reason for writing the row index as a superscript will become clear subsequently). For $a \neq b$, L^a_b is a non-Hermitian matrix with $(L^a_b)_{ba} = 1$ and all other entries zero. Note that $(L^a_b)^\dagger = L^b_a$. The $n(n - 1)$ Hermitian matrices $\frac{1}{2}(L^a_b + L^b_a)$ and $\frac{-i}{2}(L^a_b - L^b_a)$ are therefore the non-diagonal generators of SU_n. For SU_2, for example,

$$L^1_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad L^2_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{3.3}$$

are the raising and lowering matrices, and

$$L^1 \equiv \frac{L^1_2 + L^2_1}{2}, \qquad L^2 \equiv \frac{-i(L^1_2 - L^2_1)}{2} \tag{3.4}$$

are Hermitian.

For $a = b$, the Hermitian matrix L^a_a is diagonal, with $(L^a_a)^a{}_a = (n - 1)/n$ and $(L^a_a)^b{}_b = -1/n$ for $b \neq a$. The n matrices L^a_a are not independent ($\sum_a L^a_a = 0$), so a more convenient basis for the $n - 1$ diagonal generators are the matrices

$$\frac{1}{\sqrt{2k(k+1)}} \begin{pmatrix} 1 & & & & & & & 0 \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & \ddots & & & & \\ & & & & 1 & & & \\ & & & & & -k & & \\ & & & & & & 0 & \\ 0 & & & & & & & 0 \end{pmatrix} \tag{3.5}$$

for k = 1,...,n - 1. There is no particular labeling convention for the n^2-1 generators. The commutation relations are easily calculated

$$[L^a_b, L^c_d] = \delta^a_d L^c_b - \delta^c_b L^a_d \tag{3.6}$$

The abstract SU_n group is defined in terms of the n^2-1 generators T^a_b, with $(T^a_b)^\dagger = T^b_a$ and $\sum_a T^a_a = 0$, with commutation rules

$$[T^a_b,\ T^c_d] = \delta^a_d T^c_b - \delta^c_b T^a_d \tag{3.7}$$

A set of Hermitian generators T^i are defined from the T^a_b in analogy with the relation between L^i and L^a_b.

Fields transforming according to the fundamental n dimensional representation will be denoted by ψ^c, c = 1,...n, so that

$$[T^a_b,\ \psi^c] = -(L^a_b)^c{}_d\ \psi^d \tag{3.8}$$

The Hermitian conjugate $\chi_c \equiv (\psi^c)^\dagger = (\psi^\dagger)_c$ transforms according to the n^* representation

$$[T^a_b,\ \chi_c] = -\left(L^a_b(n^*)\right)_c{}^d\ \chi_d, \tag{3.9}$$

where

$$\left(L^a_b(n^*)\right)_c{}^d \equiv \left(L^a_b(n^*)\right)_{cd} \tag{3.10}$$

and

$$L^a_b(n^*) = -L^{aT}_b = -L^b_a \tag{3.11}$$

Fields transforming as n* are denoted by lower indices. Higher representations transform as direct products of n and n* representations. For example, the n^2-1 dimensional adjoint representation φ^c_d (with $\sum_a \varphi^a_a = 0$) transforms as

$$[T^a_b,\ \varphi^c_d] = -(L^a_b)^c{}_{c'}\ \varphi^{c'}_d - \left(L^a_b(n^*)\right)_d{}^{d'}\ \varphi^c_{d'} = -\delta^c_b\varphi^a_d + \delta^a_d\varphi^c_b\ . \tag{3.12}$$

Rules for the construction of irreducible representations by symmetrizing and antisymmetrizing indices, taking direct products, counting the dimensions, working with Young tableaux, etc., are described in [2.9]. It should be noted that the n* transforms like the n - 1 fold antisymmetric product of n's:

$$\chi_a \sim \varepsilon_{ab_1 \ldots b_{n-1}} \varphi^{b_1 b_2 \ldots b_{n-1}} \tag{3.13}$$

where ε is the totally antisymmetric tensor with n indices ($\varepsilon_{1,2,\ldots n} = +1$).

It is easily verified that under finite transformations,

$$\begin{aligned} \psi^c &\to U^c{}_d \psi^d = (U\psi)^d \\ \chi_c &\to \chi_d (U^\dagger)^d{}_c = (\chi U^\dagger)_c \\ \varphi^a_b &\to U^a{}_{a'} \varphi^{a'}_{b'} (U^\dagger)^{b'}{}_b = (U\varphi U^\dagger)^a{}_b , \end{aligned} \tag{3.14}$$

where U is defined in (3.1). Hence, quantities formed by contracting indices, such as

$$(\psi^c)^\dagger \psi^c = (\psi^\dagger)_c \psi^c , \tag{3.15}$$

are SU_n invariants.

For an SU_n gauge theory there are n^2-1 Hermitian gauge fields A^i. It is convenient to define the n × n matrix A by

$$\frac{1}{\sqrt{2}} A \equiv \sum_{i=1}^{n^2-1} L^i A^i = \frac{\vec{\lambda} \cdot \vec{A}}{2} , \tag{3.16}$$

where the fields

$$A^a_b \equiv (A)_{ab} \tag{3.17}$$

are non-Hermitian for a ≠ b. For the case n = 2, for example,

$$A = \begin{pmatrix} \frac{A^3}{\sqrt{2}} & A_2^1 \\ A_1^2 & \frac{-A^3}{\sqrt{2}} \end{pmatrix} , \tag{3.18}$$

where $A_2^1 = (A^1 - iA^2)/\sqrt{2}$, and where the diagonal terms are written in the basis in (3.5). For n = 3,

$$A = \begin{pmatrix} \frac{A^3}{\sqrt{2}} + \frac{A^8}{\sqrt{6}} & A_2^1 & A_3^1 \\ A_1^2 & \frac{-A^3}{\sqrt{2}} + \frac{A^8}{\sqrt{6}} & A_3^2 \\ A_1^3 & A_2^3 & \frac{-2A^8}{\sqrt{6}} \end{pmatrix} , \tag{3.19}$$

where $A_2^1 = (A^1 - iA^2)/\sqrt{2}$, $A_3^1 = (A^4 - iA^5)/\sqrt{2}$, and $A_3^2 = (A^6 - iA^7)/\sqrt{2}$. Note that

$$\mathrm{Tr}A^2 = \sum_{i=1}^{n^2-1} (A^i)^2 .$$

The covariant derivatives for the n and n* representations are

$$(D_\mu\psi)^a = [\partial_\mu\delta_b^a - ig(\vec{A}_\mu \cdot \vec{L})^a{}_b]\ \psi^b = [\partial_\mu\delta_b^a - \frac{ig}{\sqrt{2}} (A_\mu)_b^a]\ \psi^b \tag{3.20}$$

$$(D_\mu\chi)_a = \left\{\partial_\mu\delta_a^b - ig[\vec{A}_\mu \cdot \vec{L}(n^*)]_a{}^b\right\}\ \chi_b = \left\{\partial_\mu\delta_a^b + \frac{ig}{\sqrt{2}} (A_\mu)_a^b\right\}\ \chi_b \tag{3.21}$$

For the n(n - 1)/2 dimensional antisymmetric representation $\psi^{ab} = -\psi^{ba}$,

$$\begin{aligned}(D_\mu\psi)^{ab} &= \partial_\mu\psi^{ab} - ig(\vec{A}_\mu \cdot \vec{L})^a{}_c\ \psi^{cb} - ig(\vec{A}_\mu \cdot \vec{L})^b{}_d\ \psi^{ad} \\ &= \partial_\mu\psi^{ab} - \frac{ig}{\sqrt{2}} (A_\mu)_c^a\ \psi^{cb} - \frac{ig}{\sqrt{2}} (A_\mu)_d^b\ \psi^{ad}\end{aligned} \tag{3.22}$$

The gauge invariant kinetic energy term can be constructed as in (2.68). One has

$$\mathscr{L}_{kin} = -\frac{1}{4} F^{i}_{\mu\nu} F^{i\mu\nu} = -\frac{1}{4} (F_{\mu\nu})^{a}_{b} (F^{\mu\nu})^{b}_{a} \tag{3.23}$$

where

$$(F_{\mu\nu})^{a}_{b} = \partial_{\mu}(A_{\nu})^{a}_{b} - \frac{ig}{\sqrt{2}} (A_{\mu})^{a}_{c} (A_{\nu})^{c}_{b} - (\mu \leftrightarrow \nu) \tag{3.24}$$

3.3 The Georgi-Glashow Model [3.1]

The Georgi-Glashow (GG) SU_5 model [3.2] was the first attempt to embed the standard model in an underlying simple group. The model is the simplest grand unified theory that is phenomenologically viable. Many, but by no means all, of the shortcomings of the $SU^c_3 \times SU_2 \times U_1$ standard model are resolved in the GG model. I will describe the structure of the model in Section 3.3.1, relying heavily on the detailed study by Buras, Ellis, Gaillard, and Nanopoulos (BEGN) [3.3]. The strengths and weaknesses of the model are outlined in Section 3.3.2. A detailed discussion of proton decay, the baryon asymmetry of the universe, and various other theoretical and phenomenological issues is reserved for Chapters 4 through 6.

3.3.1 Basic Structure

The Fermions and Gauge Bosons

The twenty-four generators of SU_5 are T^{b}_{a}, $a,b = 1,\ldots,5$, with $T^{a}_{a} = 0$ [3.3a]. The SU^c_3 and SU_2 subgroups are generated by

$$T^{\beta}_{\alpha} - \frac{1}{3} \delta^{\beta}_{\alpha} T^{\gamma}_{\gamma} , \qquad \alpha,\beta,\gamma = 1,2,3 \tag{3.25}$$

and

$$T^s_r - \frac{1}{2}\,\delta^s_r\, T^t_t\,, \qquad r,s,t = 4,5, \tag{3.26}$$

respectively, where (α,β,γ), (r,s,t), and $(a,b,c,)$ will be used to denote SU^c_3, SU_2, and SU_5 indices, respectively.

The U_1 and electric charge generators are

$$\begin{aligned} Y &= -\frac{1}{3}T^\alpha_\alpha + \frac{1}{2}\,T^r_r \\ Q &= T^3 + Y = \frac{1}{2}(T^4_4 - T^5_5) + Y = -\frac{1}{3}T^\alpha_\alpha + T^4_4\,, \end{aligned} \tag{3.27}$$

respectively. In addition, SU_5 has twelve new generators T^r_α and T^α_r, $\alpha = 1,2,3$, and $r = 4,5$, which relate the flavor and color quantum numbers, rotate quarks into leptons, etc.

The 24 gauge bosons A^a_b transform according to the adjoint representation, which decomposes into

$$\begin{array}{ccccccccc} 24 = (8,1,0) & + & (1,3,0) & + & (1,1,0) & + & (3,2^*,-\frac{5}{6}) & + & (3^*,2,+\frac{5}{6}) \\ G^\alpha_\beta & & W^\pm, W^0 & & B & & A^\alpha_r & & A^r_\alpha \end{array}\,, \tag{3.28}$$

where the entries (n_3,n_2,y) represent the representations under the SU_3 and SU_2 subgroups and the weak hypercharge. Baryon number violating interactions are mediated by the twelve new bosons A^α_r and A^r_α, which carry both flavor and color. The color anti-triplet bosons A^4_α and A^5_α, which transform as an SU_2 doublet, are written

$$\begin{aligned} A^4_\alpha &= X_\alpha\,, \qquad Q_X = \frac{4}{3} \\ A^5_\alpha &= Y_\alpha\,, \qquad Q_Y = \frac{1}{3}\,. \end{aligned} \tag{3.29}$$

Their antiparticles are $\overline{X}^\alpha \equiv A^\alpha_4$ and $\overline{Y}^\alpha \equiv A^\alpha_5$. It is convenient to display the 5×5 matrix (3.16) $A = \sum_{i=1}^{24} A^i\lambda^i/\sqrt{2}$. It is

$$A = \left(\begin{array}{ccc|cc} G^1_1 - \frac{2B}{\sqrt{30}} & G^1_2 & G^1_3 & \bar{X}^1 & \bar{Y}^1 \\ G^2_1 & G^2_2 - \frac{2B}{\sqrt{30}} & G^2_3 & \bar{X}^2 & \bar{Y}^2 \\ G^3_1 & G^3_2 & G^3_3 - \frac{2B}{\sqrt{30}} & \bar{X}^3 & \bar{Y}^3 \\ \hline X_1 & X_2 & X_3 & \frac{W^3}{\sqrt{2}} + \frac{3B}{\sqrt{30}} & W^+ \\ Y_1 & Y_2 & Y_3 & W^- & -\frac{W^3}{\sqrt{2}} + \frac{3B}{\sqrt{30}} \end{array}\right), \quad (3.30)$$

where $G^1_1 = \frac{G^3}{\sqrt{2}} + \frac{G^8}{\sqrt{6}}$, $G^2_2 = -\frac{G^3}{\sqrt{2}} + \frac{G^8}{\sqrt{6}}$, $G^3_3 = -\frac{2}{\sqrt{6}} G^8$, and $W^\pm = (W^1 \mp iW^2)/\sqrt{2}$. I have written the diagonal elements of A in terms of the bosons in the $SU^c_3 \times SU_2 \times U_1$ subgroup for later convenience, rather than using the basis (3.5). The coefficients of the U_1 boson B are just $\sqrt{2}\, L_Y$, where L_Y is the matrix representation of Y normalized so that $\mathrm{Tr} L^2_Y = \frac{1}{2}$.

It is convenient to specify the transformation properties of the left-handed fermions and anti-fermions, as described in Section 2.4.2. Each family of fifteen fields is placed in a reducible 5* + 10 dimensional representation. The 5* field $(\psi_L)_\alpha$ has an $SU^c_3 \times SU_2 \times U_1$ decomposition

$$\begin{array}{ccccc} 5^* & = & (3^*,1,\frac{1}{3}) & + & (1,2^*,-\frac{1}{2}), \\ (\psi_L)_\alpha & & (\psi_L)_a & & (\psi_L)_r \end{array} \quad (3.31)$$

while the 10 (an antisymmetric product of two 5's) dimensional field $\psi^{ab}_L = -\psi^{ba}_L$ decomposes as

$$10 = (3^*,1,-\tfrac{2}{3}) + (3,2,\tfrac{1}{6}) + (1,1,1) \qquad (3.31)$$
$$\psi_L^{ab} \qquad \psi_L^{\alpha\beta} \qquad \psi_L^{\alpha r} \qquad \psi_L^{45}$$

Hence, ignoring for now the question of Cabibbo-like mixing between families, one can identify

$$5^*: \quad \psi_{La} = \begin{pmatrix} d_1^c \\ d_2^c \\ d_3^c \\ e^- \\ -\nu_e \end{pmatrix}_L \qquad (3.33)$$

$$10: \quad \psi_L^{ab} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & u_3^c & -u_2^c & -u^1 & -d^1 \\ -u_3^c & 0 & u_1^c & -u^2 & -d^2 \\ u_2^c & -u_1^c & 0 & -u^3 & -d^3 \\ u^1 & u^2 & u^3 & 0 & -e^+ \\ d^1 & d^2 & d^3 & e^+ & 0 \end{pmatrix}. \qquad (3.34)$$

Note that quarks, leptons, and anti-quarks (u^c and d^c) appear together in representations. The charge assignments of the fields can be verified using $[Q,\psi] = -q\psi$. The $1/\sqrt{2}$ in (3.34) is a convenient normalization factor. The identification of $(e^- \ -\nu_e)_L^T$ with a 2* field $(\psi_L)_r$ follows from $(\psi_L)_r = \varepsilon_{rs}\, \ell_L^s$, where $\ell_L = (\nu_e \ e^-)_L^T$ is a 2 and $\varepsilon_{45} = -\varepsilon_{54} = 1$. Similarly, $\psi^{54} = -\psi^{45} = \frac{1}{2}\varepsilon_{rs}\, \psi^{sr} \equiv \frac{1}{\sqrt{2}}\varphi_r^r$ is an SU_2 singlet identified with $\frac{1}{\sqrt{2}} e_L^+$. Finally,

$\psi_L^{\alpha\beta} = \frac{1}{\sqrt{2}} \varepsilon^{\alpha\beta\gamma} u_\gamma^c$ transforms like a 3* of SU_3. Equations (3.33) and (3.34) are really the fields in an interaction basis, although I have suppressed the superscripts for clarity (cf. (2.132)). Many of the signs are conventions. These signs, as well as the mixings between families, are fixed when the fields are reexpressed in the mass eigenstate basis.

It is convenient to display the fermion assignments (for three families) as

$$\begin{array}{cc}
5^* & 10 \\
\begin{pmatrix} \nu_e & \\ & d_\alpha^c \\ e^- & \end{pmatrix}_L & \begin{pmatrix} & u^\alpha & \\ e^+ & & u_\alpha^c \\ & d^\alpha & \end{pmatrix}_L \\
\begin{pmatrix} \nu_\mu & \\ & s_\alpha^c \\ \mu^- & \end{pmatrix}_L & \begin{pmatrix} & c^\alpha & \\ \mu^+ & & c_\alpha^c \\ & s^\alpha & \end{pmatrix}_L . \\
\begin{pmatrix} \nu_\tau & \\ & b_\alpha^c \\ \tau^- & \end{pmatrix}_L & \begin{pmatrix} & t^\alpha & \\ \tau^+ & & t_\alpha^c \\ & b^\alpha & \end{pmatrix}_L
\end{array} \tag{3.35}$$

where I am still neglecting mixing and am assuming that the lightest quarks are associated with the lightest leptons. (This is not a priori obvious. It must be derived by diagonalizing the fermion mass matrix.) In (3.35), pairs of fields that are associated in an SU_2 doublet are arranged in a column. SU_3^c transitions involve the color indices α. SU_5 currents coupled to the X and Y bosons involve transitions between adjacent columns. It is the presence of u, d, and u^c quarks together in the same representation that leads to fermion-number violation and proton decay in the model.

It is sometimes useful to display the right-handed charge conjugate fields

$$(\psi_R^c)^a \equiv C\,\overline{\psi}_{La}^T = \begin{pmatrix} d^1 \\ d^2 \\ d^3 \\ e^+ \\ -\nu^c \end{pmatrix}_R , \tag{3.36}$$

which transform as a 5 because of the Hermitian conjugation in (3.36).

The gauge covariant kinetic energy terms for the fermions are given by equations (3.20) and (3.22):

$$\begin{aligned} \mathscr{L}_f &= i(\overline{\psi}_R^c)_a\,(\not{D}\psi_R^c)^a + i(\overline{\psi}_L)_{ac}\,(\not{D}\psi_L)^{ac} \\ &= (\overline{\psi}_R^c)_a\,[i\,\not{\partial}\,\delta_b^a + \frac{g_5}{\sqrt{2}}\,\not{A}_b^a](\psi_R^c)^b \\ &\quad + (\overline{\psi}_L)_{ac}\,[i\,\not{\partial}\,\delta_b^a + \frac{2g_5}{\sqrt{2}}\,\not{A}_b^a]\,\psi_L^{bc} , \end{aligned} \tag{3.37}$$

where the antisymmetry of ψ_L^{bc} has been used and g_5 is the SU_5 gauge coupling. Of course, one would get the same result using (3.21) for ψ_{La} rather than (3.20) for $(\psi_R^c)^a$. The gauge part of $\mathscr{L}_f$ may be rewritten in a more convenient form, using

$$\psi_L^{\alpha\beta} = \frac{1}{\sqrt{2}}\,\varepsilon^{\alpha\beta\gamma}\,u_{L\gamma}^c$$

and the identity

$$\overline{\psi}_R^c\,\gamma_\mu\,\chi_R^c = -\overline{\chi}_L\,\gamma_\mu\,\psi_L , \tag{3.38}$$

to yield

$$\begin{aligned}\mathscr{L}_G = {} & g_5 \sum_{i=1}^{8} \left[\bar{u}\, \not{G}^i \frac{\lambda^i}{2} u + \bar{d}\, \not{G}^i \frac{\lambda^i}{2} d\right] \\ & + g_5 \sum_{i=1}^{3} \left[(\bar{u}\,\bar{d})_L \not{W}^i \frac{\tau^i}{2} \begin{pmatrix} u \\ d \end{pmatrix}_L + (\bar{\nu}_e\ \bar{e}^-)_L \not{W}^i \frac{\tau^i}{2} \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L\right] \\ & + \sqrt{\frac{3}{5}}\, g_5 \left[-\frac{1}{2}\left(\bar{\nu}_L \not{B} \nu_L + \bar{e}_L \not{B} e_L\right) + \frac{1}{6}\left(\bar{u}_L \not{B} u_L + \bar{d}_L \not{B} d_L\right)\right. \\ & \left. + \frac{2}{3}\bar{u}_R \not{B} u_R - \frac{1}{3}\bar{d}_R \not{B} d_R - \bar{e}_R \not{B} e_R\right] \\ & + \left\{\frac{g_5}{\sqrt{2}} \bar{X}^\alpha_\mu \left[\bar{d}_{R\alpha} \gamma^\mu e^+_R + \bar{d}_{L\alpha} \gamma^\mu e^+_L + \varepsilon_{\alpha\beta\gamma}\, \bar{u}^{c\gamma}_L \gamma^\mu u_L^\beta\right]\right. \\ & \left. + \frac{g_5}{\sqrt{2}} \bar{Y}^\alpha_\mu \left[-\bar{d}_{R\alpha} \gamma^\mu \nu^c_R - \bar{u}_{L\alpha} \gamma^\mu e^+_L + \varepsilon_{\alpha\beta\gamma}\, \bar{u}^{c\gamma}_L \gamma^\mu d^\beta_L\right] + \text{H.C.}\right\}\end{aligned} \tag{3.39}$$

The first two terms are just the conventional SU_3 and SU_2 interactions, provided we identify g_s and g with g_5 (at momentum scales sufficiently large that SSB can be ignored). The third term is the U_1 interaction; we must take $g' = \sqrt{\frac{3}{5}}\, g_5$, the $\sqrt{\frac{3}{5}}$ being due to the fact that $\sqrt{\frac{3}{5}}\, Y$ is a properly normalized SU_5 generator. The last two terms represent the baryon and lepton number violating interactions coupling the fermions to the X and Y bosons. It should be reemphasized that the fermion fields in (3.39) are in the interaction basis. They will be related to the mass eigenstates by unitary transformations.

Some typical interaction vertices are shown in Fig. 3.2. The X and Y bosons are known as lepto-quark bosons because of the transitions from quarks into anti-leptons which they mediate. They are also referred to as diquark bosons because two quarks can annihilate into an X or Y.

Proton Decay

Diagrams which can mediate proton or bound neutron decay are shown in Figs. 3.3 and 3.4. The diagrams in Fig. 3.3 lead to $p \to e^+ \bar{q} q$, where the $\bar{q} q$ pair can form into neutral mesons such as π^0, ρ^0, ω, η, $\pi^+\pi^-$, etc. Some of the diagrams can also lead to $n \to e^+ \bar{u} d$, where $\bar{u} d$ can be π^-, ρ^-, $\pi^-\pi^0$, etc. Such decays could of course be relevant to neutrons bound into nuclei which are energetically stable with respect to ordinary β decay. The diagram in Fig. 3.4 can generate such decays as $p \to \bar{\nu}\pi^+$, $p \to \bar{\nu}\rho^+$, $n \to \bar{\nu}\pi^0$, $\bar{\nu}\rho^0$, $\bar{\nu}\omega$, $\bar{\nu}\eta$, etc. Assuming that $M_X \sim M_Y >> m_p$ (the proton mass) we expect that the diagrams in Figs. 3.3 and 3.4 can be approximated by a four fermion interaction of strength α_5/M_X^2, where $\alpha_5 = g_5^2/4\pi$. Therefore one expects

$$\tau_p \sim \frac{1}{\alpha_5^2} \frac{M_X^4}{m_p^5} \tag{3.40}$$

Taking $\alpha_5 \sim \alpha$ and requiring $\tau_p > 10^{30}$ yr (the experimental limit) this implies $M_X \gtrsim 10^{14}$ GeV, twelve orders of magnitude larger than M_W or M_Z. However, as we shall see shortly, M_X can be predicted independently from the values of g_s, g, and g' measured at low energy, and a value of $M_X \gtrsim 10^{14}$ GeV indeed emerges, suggesting that if the GG model is correct, the proton lifetime may not be much longer than the present lower limit.

Spontaneous Symmetry Breaking

It is assumed that there are three different momentum scales relevant to the Georgi-Glashow model, as illustrated in Fig. 3.5.

(a) For $Q^2 >> M_X^2$, all SSB can be ignored and SU_5 appears to be an unbroken symmetry. The coupling constants $g_3(= g_s)$, $g_2(= g)$, and $g_1(= \sqrt{\frac{5}{3}}\, g')$

of the SU_3, SU_2, and U_1 subgroups are all equal to the SU_5 coupling constant g_5 (which is of course a function of Q^2).

(b) For $Q^2 \lesssim M_X^2$, SSB can no longer be ignored. It assumed that SU_5 is spontaneously broken down to $SU_3^c \times SU_2 \times U_1$ by an adjoint (24) representation Φ_a^b ($\Phi_a^a = 0$) of Higgs fields. The $SU_3^c \times SU_2 \times U_1$ invariant components of Φ have very large VEV's, of order 10^{14} GeV. The Φ therefore generates $M_X = M_Y \gtrsim 10^{14}$ GeV, $M_W = M_Z = m_\ell = m_q = 0$. For $M_W^2 << Q^2 \lesssim M_X^2$, $SU_3^c \times SU_2 \times U_1$ is an approximately unbroken symmetry, and g_3, g_2, and g_1 evolve independently. There are assumed to be no new thresholds associated with new bosons or fermions in the mass range between M_W and M_X, so this rather uninteresting region is referred to as the desert or plateau. Of course, one could consider more complicated models with numerous thresholds, but if the GG model in its simplest form is correct, then we can expect to see little qualitatively new physics (other than proton decay!) in the foreseeable future.

(c) It is assumed that a 5-dimensional Higgs representation H^a with a much smaller VEV of order 100 GeV breaks the $SU_3^c \times SU_2 \times U_1$ symmetry down further to $SU_3^c \times U_1^{EM}$, generating $|M_X - M_Y| \sim M_W \sim M_Z \sim$ 0(100 GeV) and $m_\ell \neq 0$, $m_q \neq 0$. Hence, for $Q^2 \lesssim M_W^2$, the observed symmetry is $SU_3^c \times U_1^{EM}$. The $SU_3^c \times SU_2 \times U_1$ content of H^a is

$$\underset{H^a}{5} = \underset{H^\alpha}{(3,1,-\tfrac{1}{3})} + \underset{\varphi}{(1,3,\tfrac{1}{2})} \tag{3.41}$$

where H^α is a color triplet of boson and φ is the Higgs doublet of the GWS model. The physical H^α fields can also mediate proton decay, so they must be made very massive.

Coupling Constant Predictions

In the standard model the coupling constants g_s, g, and g' were all arbitrary. In the SU_5 model, on the other hand, the generators are all part of the same simple group, so g_s, g, and g' are all related to each other for $Q^2 \gtrsim M_X^2$.

For example,

$$\begin{aligned} \sin^2\theta_W &= \frac{g'^2}{g'^2 + g^2} = \frac{g_1^2}{g_1^2 + \frac{5}{3} g_2^2} \xrightarrow[Q^2 \gtrsim M_X^2]{} \frac{3}{8} \\ \frac{\alpha}{\alpha_s} &= \frac{e^2}{g_s^2} = \frac{g^2 \sin^2\theta_W}{g_s^2} \xrightarrow[Q^2 \gtrsim M_X^2]{} \frac{3}{8} \end{aligned} \tag{3.42}$$

A direct asymptotic calculation of $\sin^2\theta_W$ follows from the fact that

$$\sin^2\theta_W = \frac{e^2}{g^2} = \frac{\sum_a (t_a^3)^2}{\sum_a (q_a)^2} = \frac{\mathrm{Tr}(T^3)^2}{\mathrm{Tr}(Q^2)} , \tag{3.43}$$

where the sum extends over all the fields in one representation (any representation should give the same answer). This is because

$$\begin{aligned} eQ &= g_5 \hat{Q} \\ gT^3 &= g_5 \hat{T}^3 , \end{aligned} \tag{3.44}$$

where $\hat{Q}$ and $\hat{T}^3$ are properly normalized SU_5 generators (e.g., $\mathrm{Tr}L_{\hat{Q}}^2 = \frac{1}{2}$). The prediction of $\sin^2\theta_W$ is one of the most exciting features of the GG model. When the model was first studied, the value of $\sin^2\theta_W$ was not well known experimentally, and a large value seemed reasonable. Subsequently, however, Georgi, Quinn, and Weinberg [3.4] pointed out that the $\sin^2\theta_W$ measured in neutral current experiments at $Q^2 \lesssim M_W^2$ is smaller than 3/8 because of the Q^2

dependence of the coupling constants. This is fortunate since $\sin^2\theta_W$ is now determined to be [2.70] 0.229 ± 0.009 (±0.005). Similarly, α/α_s decreases with decreasing Q^2. The Q^2 dependence of these quantities can be calculated from the renormalization group equations, as described in Section 2.5.2, so from the values of $\sin^2\theta_W$ and of α/α_s measured at low Q^2 one can in principle obtain two independent estimates of M_X. If the estimates agree then the three coupling constants g_3, g_2, and g_1 all come together at the same point, providing a consistency check on the theory. The basic equations, obtained by treating the various thresholds as step functions, ignoring Higgs fields, etc., are [3.3-3.5]

$$\frac{1}{g_i^2(Q^2)} - \frac{1}{g_i^2(M_X^2)} = -\beta_i \ \ell n \frac{Q^2}{M_X^2} , \tag{3.45}$$

where

$$\begin{aligned} \beta_1 &= \frac{f}{24\pi^2} \\ \beta_2 &= -\frac{1}{16\pi^2}\left[\frac{22}{3} - \frac{2}{3} f\right] \\ \beta_3 &= -\frac{1}{16\pi^2}\left[11 - \frac{2}{3} f\right] , \end{aligned} \tag{3.46}$$

where $f = 2F = n_q$ is the number of quark (or lepton) flavors. Equation (3.46) can easily be obtained by (2.243). One then has

$$\begin{aligned} \frac{\alpha(Q^2)}{\alpha_s(Q^2)} &= \frac{3}{8}\left[1 - \frac{11\alpha}{2\pi} \ \ell n \frac{M_X^2}{Q^2}\right] \\ \sin^2\theta_W &= \frac{3}{8}\left[1 - \frac{55\alpha}{18\pi} \ \ell n \frac{M_X^2}{Q^2}\right] \\ &= \frac{1}{6} + \frac{5}{9}\,\frac{\alpha(Q^2)}{\alpha_s(Q^2)} , \end{aligned} \tag{3.47}$$

independent of f (higher order corrections do depend on f). If one takes $\alpha/\alpha_s \simeq 1/30$ at $Q^2 = 10\ \mathrm{GeV}^2$, for example, then (3.47) implies $M_X \simeq 10^{16}$ GeV and $\tau_p \gtrsim 10^{37}$ yr [3.3], too long to be observable in planned experiments. $\mathrm{Sin}^2\theta_W$ is predicted to be $\simeq 0.19$ at $Q^2 = 10\ \mathrm{GeV}^2$, which is approximately the right value. As we will see in Chapter 4, small corrections to these formulas will imply large corrections to the value of M_X and therefore to $\tau_p \sim M_X^4$. Current estimates yield $M_X \gtrsim 10^{14}$ GeV and $\tau_p \gtrsim 10^{30}$ yr, which is precisely the region to which new experiments will be sensitive!

The Higgs Sector

Let us now consider the Higgs sector of the theory in more detail. The adjoint Higgs representation Φ_a^b ($\Phi_a^a = 0$) can be written as a matrix (similar to the matrix A for the gauge bosons).

$$\Phi \equiv \sum_{i=1}^{24} \varphi^i \frac{\lambda^i}{\sqrt{2}} \quad , \tag{3.48}$$

with $\Phi_a^b = (\Phi)_{ba}$. If Φ were the only Higgs representation, then [3.3]

$$\begin{aligned} V(\Phi) = &- \frac{\mu^2}{2} \mathrm{Tr}(\Phi^2) + \frac{1}{4} a\left[\mathrm{Tr}(\Phi^2)\right]^2 \\ &+ \frac{1}{2} b\ \mathrm{Tr}(\Phi^4) + \frac{1}{3} c\ \mathrm{Tr}\Phi^3 \ . \end{aligned} \tag{3.49}$$

V is usually simplified by imposing a symmetry under $\Phi \to -\Phi$, so that $c = 0$. The VEV $\langle 0|\Phi|0\rangle$ can always be taken to be diagonal by performing a suitable SU_5 transformation (3.14). Then, according to Li [3.6] V will be minimized for $\langle 0|\Phi|0\rangle = \mathrm{diag}\ (\nu,\nu,\nu,\nu, -4\nu)$ if $b < 0$ or $\langle 0|\Phi|0\rangle = \mathrm{diag}\ (\nu,\nu,\nu, -\frac{3}{2}\nu, -\frac{3}{2}\nu)$ if $b > 0$, where diag (a,b,c, ...) refers to a diagonal matrix with entries a, b, c, ... SU_5 is spontaneously broken to $SU_4 \times U_1$ or to

$SU_3 \times SU_2 \times U_1$, for the two cases, respectively. We are interested in the latter situation, so we will assume $b > 0$ and $a > -7b/15$, which is needed for positivity [3.3]. Then [3.3]

$$\nu^2 = \frac{2\mu^2}{15a + 7b} \tag{3.50}$$

and the X and Y boson masses are

$$M_X^2 = M_Y^2 = \frac{25}{8} g_5^2 \nu^2 \tag{3.51}$$

Twelve of the Higgs fields $\Phi - \langle 0|\Phi|0\rangle$ are eaten by the Higgs mechanism and the other twelve bosons have super heavy masses of order $\sqrt{b}\,\nu$. They do not couple to fermions and are therefore of limited interest.

SU_5 can be broken down to $SU_4 \times U_1$ if $b < 0$, if $c \neq 0$, or possibly if radiative corrections are important. Guth and Tye [3.7] have considered the case $c \neq 0$ in connection with the suppression of magnetic monopole production (Chapter 6).

In the presence of two Higgs representations, Φ and H, it is possible to break the symmetry down to $SU_3^c \times U_1^{EM}$. The fundamental H^a consists of H^α, a color triplet, and an SU_2 doublet φ that plays the role of the doublet in the GWS model. If one considers a potential

$$V(H) = -\frac{\mu_5^2}{2} H^\dagger H + \frac{\lambda}{4} (H^\dagger H)^2 \quad , \tag{3.52}$$

with no cross coupling between Φ and H, then, assuming SU_3^c is not broken, one has $\langle 0|H^5|0\rangle = \frac{1}{\sqrt{2}}\nu_0$, where $\nu_0^2 = 2\mu_5^2/\lambda$ and $M_W^2 = g^2\nu_0^2/4$ (I have modified the notation from Chapter 2 to conform to that of BEGN). Unfortunately, the theory contains two neutral color triplets of Higgs fields, H^α and Φ_5^α. One linear combination, which consists mainly of Φ_5^α, is eaten to give mass to the

Y boson. However, the orthogonal combination, which is approximately H^α, represents an unwanted massless colored Higgs particle that could mediate proton decay.

The problem can be removed by adding cross terms (which are in general present)

$$V(\Phi,H) = \alpha H^\dagger H \,\mathrm{Tr}\Phi^2 + \beta H^\dagger \Phi^2 H + \delta H^\dagger \Phi H \tag{3.53}$$

to the potential to give mass to the H^α. If the $\Phi \to -\Phi$ discrete symmetry is imposed then $\delta = 0$ and the extremum of the combined potential is [3.3] $\langle 0|\Phi|0\rangle = \mathrm{diag}\,[\nu,\nu,\nu,\ (-\frac{3}{2} - \frac{1}{2}\varepsilon)\nu,\ (-\frac{3}{2} + \frac{1}{2}\varepsilon)\nu]$ and $\langle 0|H^5|0\rangle = \frac{1}{\sqrt{2}}\nu_0$ where

$$\varepsilon = \frac{3}{20}\frac{\beta\nu_0^2}{b\nu^2} + O\left(\frac{\nu_0^4}{\nu^4}\right) . \tag{3.54}$$

Note that $\varepsilon\nu << \nu_0$ for $\nu_0 << \nu$. This is good because $\Phi_4^4 - \Phi_5^5$ is an SU_2 isotriplet Higgs field, the VEV of which is strongly constrained by the neutral current data (Section 2.4.4). ν and ν_0 are determined by

$$\begin{aligned} \mu^2 &= \frac{15}{2} a\nu^2 + \frac{7}{2} b\nu^2 + \alpha\nu_0^2 + \frac{9}{30}\beta\nu_0^2 \\ \mu_5^2 &= \frac{1}{2}\lambda\nu_0^2 + 15\,\alpha\nu^2 + \frac{9}{2}\beta\nu^2 - 3\,\varepsilon\beta\nu^2 . \end{aligned} \tag{3.55}$$

Furthermore, the physical color triplet Higgs field is (for $\beta < 0$)

$$h^\alpha \equiv H^\alpha + \frac{\sqrt{2}\nu_0}{5\nu}\Phi_5^\alpha + O\left(\frac{\nu_0^2}{\nu^2}\right) , \tag{3.56}$$

with

$$\mu_h^2 = -\frac{5}{2}\beta\nu^2 + O\left(\nu_0^2\right) \tag{3.57}$$

We wish to choose parameters so that $\mu_h^2 \lesssim M_X^2$ (so that h exchange does not lead to too short a proton lifetime) and $\nu_0^2 <<< \nu^2$. This is clearly possible, but it requires a very delicate adjustment or tuning (to one part in $\nu^2/\nu_0^2 \simeq 10^{24}$) of the parameters in (3.55). This is our first glimpse of the hierarchy problem. For most values of the parameters in the Higgs potential, ν_0^2 and ν^2 (and therefore M_W^2 and M_X^2) will be of the same order of magnitude. To obtain a hierarchy of very different scales a very fine tuning of parameters appears to be necessary.

A description of the physical Higgs particles is given by BEGN [3.3]. Magg and Shafi [3.8] and Sherry [3.9] have considered the minimization of the potential $V(\Phi) + V(H) + V(\Phi,H)$ in more detail, and have found that for $a > -7b/15$, $b > 0$, $\beta < 0$, SU_5 is indeed broken down to $SU_3^c \times U_1^{EM}$ (other patterns of breaking occur for other ranges of the parameters). Buccella, Ruegg and Savoy [3.10] have extended the analysis to the breaking of SU_n by one adjoint and one fundamental Higgs representation. These analyses all impose the discrete symmetry under $\Phi \to -\Phi$. Scott [3.11] has shown that the desired breaking can also occur for a special case of the model for which this discrete symmetry has not been imposed ($c \neq 0$, $\delta \neq 0$). The most general case with $c \neq 0$, $\delta \neq 0$ has been studied (for SU_n) by Ruegg [3.12]. The desired breaking of SU_5 to $SU_3^c \times U_1^{EM}$ can occur for a range of the parameters. Symmetry breaking by several adjoint [3.13] or several antisymmetric tensor [3.14] Higgs representations has also been considered.

Fermion Masses

The left-handed fermions are assigned to 5* and 10 dimensional representations ψ_{La} and ψ_L^{ab}. The form of a mass term for two left-handed fields ψ_L and χ_L is (ignoring SU_5 indices)

$$\psi_L^T C \chi_L + \text{H.C.} = \chi_L^T C \psi_L + \text{H.C.} \equiv \overline{\psi_R^c} \chi_L + \text{H.C.} \,, \tag{3.58}$$

where $C = -C^T$ is the charge conjugation matrix defined in (2.110). Yukawa terms have the same structure, but are multiplied by a scalar field. Hence, Yukawa terms for the SU_5 model must be of the form

$$\begin{aligned} &\psi_{La}^T C \psi_{Lb} + \text{H.C.} \\ &\psi_{La}^T C \psi_L^{bc} + \text{H.C.} \\ &\psi_L^{Tab} C \psi_L^{cd} + \text{H.C.} \end{aligned} \tag{3.59}$$

with an appropriate contraction of indices with the Higgs fields. Family indices can be added if desired. The first and third terms, if present, imply the violation of fermion number by two units in the model. The second term is invariant w.r.t. fermion number if ψ_L^{ab} and ψ_{La} are assigned fermion numbers ± 1, respectively.

The decomposition of the direct products in (3.59) are

$$\begin{aligned} 5^* \times 5^* &= 10 + 15 \\ 5^* \times 10 &= 5 + 45 \\ 10 \times 10 &= 5^* + 45^* + 50 \,. \end{aligned} \tag{3.60}$$

None of these products include a 24, so the adjoint representation Φ_a^b does not couple to fermions (this is fortunate because otherwise the natural scale for fermion masses would be $m_f \lesssim M_X$). There could be couplings

$$\psi_{mLa}^T C \psi_{nLb} H_{S,A}^{ab} \tag{3.61}$$

of two 5*'s (m and n are family indices) to a symmetric (15) or antisymmetric (10) dimensional Higgs ($H_S^{ab} = H_S^{ba}$, $H_A^{ab} = -H_A^{ba}$), if these were introduced into the theory. The 10 does not contain any neutral color singlet, so it would not contribute to the fermion masses if $SU_3^c \times U_1^{EM}$ is unbroken. The 15 contains a color singlet that transforms as an isotriplet under SU_2. Note that

$$\psi^T_{mLa}\ C\ \psi_{nLb} = \psi^T_{nLb}\ C\ \psi_{mLa} \quad , \tag{3.62}$$

so that the couplings to H_S or H_A would have to be symmetric or anti-symmetric in the family indices, respectively.

If only the Φ_a^b and H^a fields are included (The effects of a 45 will be discussed below. The 50, like the 10, has no neutral color singlet component.), then the only Yukawa couplings are

$$\gamma_{mn}\ \psi^T_{mLa}\ C\ \psi^{ab}_{nL}\ H_b^\dagger + \text{H.C.} = \gamma_{mn} \left(\overline{\psi^c_{mR}}\right)_a \psi^{ab}_{nL}\ H_b^\dagger + \text{H.C.} \tag{3.63}$$

and

$$\Gamma_{mn}\ \varepsilon_{abcde}\ \psi^{Tab}_{mL}\ C\ \psi^{cd}_{nL}\ H^e + \text{H.C.} \tag{3.64}$$

where ε_{abcde} is the totally anti-symmetric tensor. It is the coupling (3.64) that violates fermion number and leads to proton decay in the GG model. The interaction is symmetric in the family indices m and n so that Γ_{mn} can be taken to be symmetric.

For $\langle 0|H^a|0\rangle = \frac{1}{\sqrt{2}}\nu_0\ \delta_a^5$, the Yukawa coupling (3.63) generates the fermion mass matrix

$$-\frac{\nu_0}{2}\gamma_{mn}\left(\overline{d}_{mR}\ d_{nL} + \overline{e^+_{mR}}\ e^+_{nL}\right) + \text{H.C.} = -\ \overline{d}_R\ M^d\ d_L - \overline{e^+_R}\ M^e\ e^+_L + \text{H.C.} \tag{3.65}$$

where

$$M^d = M^e = \frac{\nu_0}{2}\gamma \ . \tag{3.66}$$

(For notational convenience, M^d is the adjoint of the matrix defined in Section 2.4.3.) That is, the SU_5 symmetry requires that the d quark and positron mass matrices must be the same. This result continues to hold if several five dimensional Higgs fields are included. In particular, the eigenvalues are the same:

$$
\begin{aligned}
m_d &= m_e \\
m_s &= m_\mu \\
m_b &= m_\tau
\end{aligned}
\tag{3.67}
$$

At first sight the relations (3.67) appear to be a disaster. However, the fermion masses must be interpreted as effective masses with values that depend on the momentum at which they are measured. The predictions (3.67) only apply for $Q^2 \gtrsim M_X^2$. The Q^2 dependence of the fermion mass operators has been computed by BEGN [3.3, 3.5], with the result that for $Q^2 < M_X^2$,

$$
\begin{aligned}
\ell n \left[\frac{m_d(Q^2)}{m_e(Q^2)}\right] &= \ell n \left[\frac{m_d(M_X^2)}{m_e(M_X^2)}\right] \\
&+ \frac{4}{11 - 2n_q/3} \ \ell n \left[\frac{\alpha_s(Q^2)}{\alpha_5(M_X^2)}\right] + \frac{3}{2n_q} \ \ell n \left[\frac{\alpha_1(Q^2)}{\alpha_5(M_X^2)}\right] ,
\end{aligned}
\tag{3.68}
$$

with similar results holding for the other ratios. It is conventional [3.15] to define a "physical" current quark mass by the value $m_q(Q_0^2)$ where Q_0^2 is defined by $\sqrt{Q_0^2} = 2m_q(Q_0^2)$. This definition is excellent for heavy quarks, questionable for the s quark, and inadequate for the u and d quarks. Nanopoulos and Ross [3.16] have added threshold and higher order effects to (3.68). They find that m_b is predicted to be 5.3, 5.8, and 6.9 GeV to lowest order for n_q = 6, 8, or 10, respectively. These values are for Λ = 300 MeV. The effect of higher order corrections is to increase these predictions by $\gtrsim$1 GeV. These values are somewhat high compared to the expected value of $\simeq$5 GeV, but one may view this approximate agreement as a triumph for the model. Nanopoulos and Ross [3.16] and BEGN [3.3] argue that the strong n_q dependence of the predictions requires that there be no more than three families ($n_q = 6$) of fermions. For the s quark, the situation is less satisfactory. The prediction to lowest order is $m_s = 470$ MeV for $n_q = 6$. The prediction is increased by higher order corrections or for $n_q > 6$. This value seems rather high. Typical estimates of the s quark current mass are 150-300 MeV [2.168], but the theoretical uncertainty is at least a factor of two. The absolute value of the d quark current mass is even more difficult, but fortunately the ratio m_d/m_s of current algebra masses is essentially independent of renormalization effects [2.143]. Hence, (3.67) implies

$$\frac{m_d}{m_s} = \frac{m_e}{m_\mu} \simeq \frac{1}{207} \tag{3.69}$$

This is to be compared to the phenomenological value [2.143] of $m_d/m_s \simeq \frac{1}{24}$ obtained from the meson and baryon mass spectra (see Section 2.5.2). This order of magnitude discrepancy is a serious problem for the SU_5 model with the minimal Higgs structure. Some possible resolutions of the problem are considered in Chapter 6. One can also introduce a 45 dimensional Higgs multiplet H^{ab}_c ($H^{ab}_c = -H^{ba}_c$; $H^{ab}_a = 0$) to the model [3.17, 3.18], with couplings

$$\psi^T_{La}\, C\, \psi^{bc}_L\, H^{\dagger a}_{bc} + \text{H.C.} \tag{3.70}$$

and

$$\varepsilon_{abcef}\, \psi^{Tab}_L\, C\, \psi^{cd}_L\, H^{ef}_d + \text{H.C.}\ . \tag{3.71}$$

The coupling in (3.71) is antisymmetric in the family labels (unlike the coupling (3.64) which is symmetric). The $SU^c_3 \times U^{EM}_1$ invariant component of H is [3.17-3.18]

$$\langle 0|H_a^{b5}|0\rangle = \nu_{45}\ (\delta_a^b - 4\delta_4^a\ \delta_b^4) \tag{3.72}$$

for a,b = 1, ... 4. The 45 generates d and e mass matrices related by

$$M_{45}^e = -3M_{45}^d \tag{3.73}$$

If both 5 and 45 representations contribute to M^d and M^e, then

$$M^e = M_5^e + M_{45}^e = M_5^d - 3M_{45}^d\ , \tag{3.74}$$

so that in general there is no relation between the d and e masses and mixings. However, if only the 45 is present, then (3.67) is replaced by

$$\begin{aligned} m_e &= 3m_d \\ m_\mu &= 3m_s \\ m_\tau &= 3m_b\ , \end{aligned} \tag{3.75}$$

after redefining the phases of the right-handed fields. The problem with $m_d/m_s = m_e/m_\mu$ is unaffected. Frampton et al. [3.17] have argued that in this case the correct value of m_b can be obtained if $n_q = 12$ (i.e., six families). An interesting model by Georgi and Jarlskog [3.18] involving both a 5 and 45, which leads to the correct value for m_d/m_s is described in Chapter 6.

Let us now consider the mass terms generated by the 10 × 10 × 5 term in (3.64). It is

$$\begin{aligned} \frac{4\nu_0}{\sqrt{2}}\Gamma_{mn}\ \varepsilon_{\alpha\beta\gamma}\ \psi_{mL}^{T\alpha\beta}\ C\ \psi_{nL}^{\gamma 4} + \text{H.C.} &= -\frac{4\nu_0}{\sqrt{2}}\Gamma_{mn}\ \bar{u}_{mR}\ u_{nL} + \text{H.C.} \\ &= -\ \bar{u}_R\ M^u\ u_L + \text{H.C.}\ , \end{aligned} \tag{3.76}$$

where

$$M^u = \frac{4\nu_0}{\sqrt{2}}\Gamma = M^{uT} \tag{3.77}$$

is the symmetric u quark mass matrix. The coupling (3.71) of a 45 would generate an antisymmetric mass matrix. For three families a purely antisymmetric mass matrix (i.e., that generated by a 45 only) would lead to the bad results $m_u = 0$, $m_c = m_t$ [3.19, 3.20].

The total fermion mass term is therefore

$$\mathscr{L}_F = - \bar{u}^0_R M^u u^0_L - \bar{d}^0_R M^d d^0_L - \bar{e}^{0+}_R M^e e^{0+}_L + \text{H.C.} , \tag{3.78}$$

where I have added superscripts 0 to indicate the interaction basis. M^u, M^d, and M^e are arbitrary F × F matrices in general, but if the masses are all generated by an arbitrary number of 5 dimensional Higgs representations then because of the SU_5 symmetry we have the restrictions $M^u = M^{uT}$ and $M^d = M^e$. The positron term in (3.78) could be rewritten in terms of electron fields as $-\bar{e}^0_R M^{eT} e^0_L$ if desired. There are no neutrino mass terms in (3.78), but they can be added if desired by introducing SU_5 singlet ν_R fields which couple to ψ_{La} via the H^a, just as in $SU_2 \times U_1$ (see Section 2.4.4).

The u, d, and e^+ mass matrices can be diagonalized in the same way as in the GWS model. One defines $d^0_{L,R} = A^d_{L,R}\, d_{L,R}$ such that

$$A^{d\dagger}_R M^d A^d_L = M^d_D = \begin{pmatrix} m_d & & & 0 \\ & m_s & & \\ & & m_b & \\ 0 & & & \ddots \end{pmatrix} \tag{3.79}$$

with similar definitions for $A^{e+}_{L,R}$ and $A^u_{L,R}$. The $A_{L,R}$ are determined up to phase matrices $K_{L,R}$ by the condition that M^2_D be real or diagonal; the relative phases $K_L K^*_R$ are determined by the reality and positivity of M_D, and the individual phases K_L are unobservable but may be chosen to put the matrices into a convenient standard form. In the GWS model the mixing matrices A_R^u, A_R^{d-}, and, in the present notation, A_L^{e+}, were not observable because the fields did not participate in non-diagonal interactions. In the SU_5 model, however, these matrices are observable in the lepto-quark and diquark interactions.

It remains to express the interaction fields d_{mL}^{oc} and u_{mL}^{oc} in terms of the mass eigenstates d_{nR} and u_{nR}. Recall that

$$d_{mL}^{oc} \equiv C \left(\overline{d_{mR}^{o}}\right)^T \tag{3.80}$$

and that

$$d_{mR}^{o} \equiv \left(A_{L,R}^{d}\right)_{mn} d_{nR} \tag{3.81}$$

Hence,

$$d_{mL}^{oc} = \left(A_R^{d\dagger}\right)_{nm} C \left(\overline{d_{nR}}\right)^T \equiv d_{nL}^{c} \left(A_R^{d\dagger}\right)_{nm} ; \tag{3.82}$$

where d_{nL}^{c} is <u>defined</u> as $C\left(\overline{d_{nR}}\right)^T$ (i.e., it is <u>not</u> the same as $\left(A_L^{d\dagger}\right)_{nm} d_{mL}^{oc}$). Finally, we are free to pick basis states such that e_L^- and d_L are diagonal; then the 5* and 10 representations, in terms of mass eigenstates, are (in the notation of (3.35))

$$\begin{aligned} &5^* \begin{pmatrix} \nu_m & \\ & \left(d^c A_R^{d\dagger}\right)_m \\ e_m^- & \end{pmatrix}_L \\ &10 \begin{pmatrix} & \left(A_L^u u\right)_m & \\ \left(A_L^{e^+} e^+\right)_m & & \left(u^c A_R^{u\dagger}\right)_m \\ & d_m & \end{pmatrix}_L . \end{aligned} \tag{3.83}$$

2F-1 of the 3F phases in the K_L^d, K_L^u, and $K_L^{e^+}$ matrices may be chosen to put the KM matrix A_L^u into the standard $(F-1)^2$ parameter form. F phases may be chosen to simplify $A_L^{e^+}$, which therefore has F^2-F observable parameters. The final phase may be chosen to simplify A_R^u so that A_R^u and A_R^d have F^2-1 and F^2 observable parameters, respectively.

In the general case, the matrices $A_{L,R}$ are arbitrary. It would even be possible to choose mass matrices so that, for example, the u and d quarks are associated in multiplets with the τ lepton, which would greatly suppress proton decay [3.21]. However, the situation simplifies enormously if only 5 dimensional Higgs are included [3.3,3.22-23]. One then has $A_{L,R}^d = A_{L,R}^{e^+} = I$ (in the basis being used). Moreover, the symmetry $M^u = M^{uT}$ implies (for a given A_L^u)

$$A_R^{u*} = A_L^u \; K^* \tag{3.84}$$

where K is a diagonal matrix of phases (assuming no degeneracy of the eigenvalues) which is uniquely determined by the condition that M_D^u be real and positive. Then the 5* and 10 fields are

$$\begin{aligned} &5^* \begin{pmatrix} \nu_m & \\ & d_m^c \\ e_m^- & \end{pmatrix}_L \\ &10 \begin{pmatrix} & \left(A_C\right)_{mn} u_n & \\ e_m^+ & & \left(A_C\right)_{mn} e^{-i\alpha_n} u_n^c \\ & d_m & \end{pmatrix}_L \end{aligned} \tag{3.85}$$

where A_C is the generalized Cabibbo matrix and $\exp(-i\alpha_n)$ is the n^{th} diagonal entry of K^* (only F-1 of these phases are observable. The last corresponds to an arbitrary phase of all of the fields in the theory). Hence, except for the extra phases all of the mixing matrices are determined by the Cabibbo matrix;—the light quarks are associated with the light leptons, and proton decay cannot be "rotated away" [3.3,3.22-23].

The interactions between the X and Y bosons and fermions can be obtained by rewriting (3.39) in terms of the mass eigenstates. In the special case (3.85) one has

$$\begin{aligned} &\frac{g_5}{\sqrt{2}} \bar{X}^{\alpha}_{\mu} \left[\bar{d}_{R\alpha} \gamma^{\mu} e^{+}_{R} + \bar{d}_{L\alpha} \gamma^{\mu} e^{+}_{L} + \varepsilon_{\alpha\beta\gamma} \overline{u}^{c\gamma}_{L} K \gamma^{\mu} u^{\beta}_{L}\right] \\ &+ \frac{g_5}{\sqrt{2}} \bar{Y}^{\alpha}_{\mu} \left[- \bar{d}_{R\alpha} \gamma^{\mu} \nu^{c}_{R} - \bar{u}_{L\alpha} A^{\dagger}_{C} \gamma^{\mu} e^{+}_{L} + \varepsilon_{\alpha\beta\gamma} \overline{u}^{c\gamma}_{L} K A^{\dagger}_{C} \gamma^{\mu} d^{\beta}_{L}\right] \quad (3.86) \\ &+ \text{H.C.} \ , \end{aligned}$$

where u, d, e^+, and ν are F component vectors in family space.

3.3.2 Features of the Model

In this section I summarize the major strong and weak points of the Georgi-Glashow model and comment on analogous features in more elaborate models. Many of these topics will be discussed in more detail in Chapters 4 through 6.

Among the attractive features of the SU_5 model are the following:

(a) The SU_5 model incorporates the $SU^c_3 \times SU_2 \times U_1$ model as a maximal subgroup. It is the only acceptable theory with a single coupling constant with this property [3.2]. Many larger groups also incorporate the standard model but there is generally more freedom in the pattern of SSB.

(b) The structure of the charged weak current, including maximal parity violation and generalized Cabibbo universality of the quark and lepton currents are natural features of the model for the $5^* + 10$ assignment of the fermions. Larger groups or different representations within SU_5 do not always have these features.

(c) There is no room for a ν_R in the $5^* + 10$ representations. Hence, the neutrino is massless unless an SU_5 singlet ν_R or a 15 dimensional Higgs field (which allows a Majorana mass for ν_L; this is just the Higgs triplet model of Chapter 2) are introduced. Most larger groups include ν_R fields so that the neutrino will generally be massive.

(d) Electric charge is quantized (i.e., the quark and lepton charges are related) because the electric charge operator is an SU_5 generator. Therefore $\mathrm{Tr}L_Q = 0$ and the sum of the charges of the particles in each multiplet must be zero. For the 5^*, for example, $3\, q_{d^c} + q_{e^-} = 0$ or $q_d = \frac{1}{3} q_{e^-} = -\frac{1}{3}$. The factor of 3 is the number of colors [3.24]: for an SU_n color group the generalization of the model to SU_{n+2} would yield $q_d = q_{e^-}/n$. For larger groups the quark and lepton charges may or may not be related, depending on the details of the model [3.25].

(e) The fact that $\alpha_s \gg \alpha$ at low energies implies a very large unification mass, as is required by the approximate stability of the proton. Most other models share this feature.

(f) The proton and bound neutrons are predicted to decay with an experimentally accessible lifetime of $\tau_p \gtrsim 10^{30}$ yr. Baryon and lepton number are both violated, but the combination B-L is conserved. Most other models also lead to proton decay but the lifetime and branching ratios depend on the model.

(g) $\sin^2\theta_W$ is predicted to be $\simeq 0.20$-0.21, in the right general range but slightly lower than the present experimental values. The prediction of $\sin^2\theta_W$ and the approximate validity of B-L are shared by all models in which there is a desert between M_W and $M_X \gtrsim 10^{14}$ GeV in which $SU_3^c \times SU_2 \times U_1$ is approximately unbroken and there are no thresholds. Models with several stages of symmetry breaking may allow different values for $\sin^2\theta_W$ and substantial B-L violation.

(h) The baryon number violating interactions may be able to explain the asymmetry between baryons and anti-baryons in the universe, with the important consequence that we can exist. It appears that the ratio $n_B/n_\gamma \sim 10^{-9\pm1}$ can be generated in the model, but this depends on the number of Higgs multiplets (more than one 5 are required), on the origin of CP violation, and on the details of the cosmology. Most other models are similar.

(i) The simplest Higgs scheme, with only 24 and 5 dimensional representations, gives an approximately correct prediction of $m_b/m_\tau \gtrsim 3$ for three families. The prediction would fail for more families. Most other models have more freedom.

(j) There are no flavor changing neutral current effects associated with the light gauge bosons. If more than one Higgs multiplet couple to fermions, as is suggested by the baryon asymmetry constraints and possibly by the m_d/m_s ratio, then unless extra symmetries are imposed there will in general be FCNC effects mediated by the Higgs particles. Their strength depends on the Higgs masses, just as in the GWS model. Some of the more complicated models include horizontal gauge symmetries which can also mediate FCNC processes.

The less attractive features of the model include:

(a) Each family is placed in a reducible 5* + 10 representation. In larger groups each family is usually in an irreducible representation. In the SO_{10} model, for example, each family is placed in a 16 dimensional representation, which consists of the 5* + 10 of the SU_5 subgroup, as well as an SU_5 singlet, which is the ν_R.

(b) There is no explanation of why there are several families or of how many there are (except for the phenomenological constraint on m_b/m_τ and the "need" to have $\gtrsim 3$ families to have CP violation in the ordinary weak interactions). The SU_5 model has made no progress with respect to the standard model on this particular question. Larger groups often include horizontal interactions that restrict the family structure. In some cases the families are included in a single irreducible representation or in a direct sum of irreducible representations without any repetition.

(c) The difficulties with the predictions for m_s and m_d/m_s have already been described. Possible solutions include introducing more complicated Higgs representations, such as a 45 [3.18], or postulating that there may be small (a few MeV) corrections to the fermion mass matrix from other sources, such as effective nonrenormalizable interactions generated somehow by quantum gravity [3.26]. Also, there are no constraints on the u, c, and t masses (except that the mass matrix is symmetric) unless additional symmetries are imposed on the Yukawa couplings [3.27]. In larger groups not involving horizontal gauge symmetries the situation is usually similar: a particular Higgs representation may generate a symmetric or antisymmetric mass matrix and may relate the quark and lepton masses. It will generally not relate the masses in different families or determine the mixing angles unless additional symmetries are imposed. The mass relations determined then are as much a function of the extra symmetries as of the gauge symmetry. Much stronger restrictions should in principle come about in models with horizontal gauge symmetries, but in practice the results will depend on the details of the Higgs representations and the pattern of spontaneous symmetry breaking. Few examples have been worked out in detail.

(d) The model still has many free parameters. The model with only one 24 and one 5 Higgs representation has [2.60] 1 gauge coupling, 1 θ parameter, 9 Higgs parameters (7 if the discrete $\Phi \to -\Phi$ symmetry is imposed), 6 quark masses (from which the lepton masses are deduced), and 6 mixing angles and CP violating phases, for a total of 23. I have not subtracted an overall mass scale because the super heavy masses are close to the Planck mass, and I have not included a primordial cosmological term. This is to be compared with 19 (or 20) found in the standard model without right-handed neutrinos. The Higgs sectors of most larger groups have not been considered in enough detail to count the free parameters.

(e) The SU_5 model allows the existence of super heavy ($m \simeq 10^{16}$ GeV) 't Hooft-Polyakov magnetic monopoles [3.28], which are topologically stable classical configurations of the gauge and Higgs fields of a spontaneously broken gauge theory [3.29]. These may have been produced in unacceptably large numbers in the early universe unless some mechanism is invoked to suppress their production or to enhance the annihilation of monopole-antimonopole pairs (Chapter 6). The situation is similar in most other theories.

(f) The model includes a desert between the W and X masses. This is not necessarily a problem, but it would be very boring. Of course one can introduce additional thresholds in this region, but one then tends to lose predictive power.

(g) Closely related to the desert is the hierarchy problem. The GG model requires two distinct mass scales in the ratio $M_W^2/M_X^2 \simeq 10^{-24}$. The existence of two scales is not a natural feature of the model. It is necessary to adjust or fine tune the parameters in the Higgs potential, including the loop corrections [3.30], to one part in 10^{24} to achieve this hierarchy. The hierarchy problem exists in most other models unless they manage to have a small unification mass or to have a whole sequence of closely spaced thresholds betweem M_W and M_X. Possible explanations of the hierarchy problem are discussed in Chapter 6.

(h) Grand unified models do not include gravity. Supersymmetric theories [3.31] are a very promising approach to this problem, but no completely satisfactory examples have been given as of yet.

Other issues, such as CP violation, asymptotic freedom, and neutral currents, are discussed in Chapter 6.

3.4 Other Models

There have been many suggestions that SU_5 may be a subgroup of a still larger gauge theory. Among the possible roles of the new generators of a larger group are the following:

(a) They could connect the elements of the 5* and 10 representations, so that each family of fermions would be contained in an irreducible representation of the larger group.

(b) They could connect the families (i.e., generate a horizontal symmetry).

(c) They could generate an entirely new type of interaction, such as technicolor or extended technicolor.

In this section the basic constraints on extensions of the SU_5 model and on alternatives will be described, along with a number of examples that have been proposed. General theoretical and phenomenological issues will be considered in Chapters 4 through 6.

3.4.1 General Constraints

There are a number of general constraints that are often imposed on grand unified theories. These constraints should be viewed as useful guidelines and not as inviolate rules.

Single Coupling Constant

One of the motivations for considering GUTS is to reduce the number of free parameters. One desirable constraint is to demand that the theory have only one gauge coupling constant (this can be taken to be the definition of a grand unified theory). In order to have only one gauge coupling, the gauge group G must either be a simple group such as SU_n (simple groups are defined and classified below) or a direct product of identical simple groups, such as $SU_n \times SU_n \times SU_n$. In the latter case some sort of refection symmetry that interchanges the factors must be imposed on the theory to force the (normally independent) gauge couplings to be equal. This in turn requires that there be a one-to-one correspondence of the fermion and Higgs representations of the factors. Direct products of nonidentical factors cannot in general have the same gauge coupling for the factors. Even if one artificially sets the renormalized running couplings equal at some momentum they will in general be different at other momentum scales. (However, Levin [3.32] has found examples of products of two nonidentical factors for which a relation between the gauge couplings can be preserved up to the two loop level by a judicious choice of fermion representations.)

Classification of Simple Lie Groups [2.9]

A subgroup H of a Lie group G is an invariant subgroup if $ghg^{-1} \varepsilon H$ for all $g \varepsilon G$, $h \varepsilon H$. G is called a simple group if it contains no invariant subgroup (other than the identity and G itself). G is called semi-simple if it contains no abelian invariant subgroup. Compact semi-simple groups are either simple or the direct product of two or more simple groups. For example, SU_n is simple (and semi-simple), $SU_n \times SU_m$ is semi-simple, and $SU_2 \times U_1$ is neither simple nor semi-simple.

The algebras of the generators of the simple Lie groups have been classified by Cartan. There are four countable sequences of simple Lie algebras, A_ℓ, B_ℓ, C_ℓ, and D_ℓ, where the rank ℓ is a positive integer. These are identified with the generators of the classical groups $SU_{\ell+1}$, $SO_{2\ell+1}$, $Sp_{2\ell}$, and $SO_{2\ell}$, respectively. These groups are defined in terms of their fundamental representations. SU_n are the $n \times n$ complex unitary matrices of unit determinant. They leave invariant the inner product of two vectors in an n dimensional complex vector space. SO_n are the $n \times n$ real orthogonal matrices with unit determinant. They leave inner products in an n dimensional real vector space invariant. The symplectic matrices Sp_{2n} are real $2n \times 2n$ matrices M which leave invariant the skew symmetric matrix

$$S = \begin{pmatrix} 0 & 1 & & & & 0 \\ -1 & 0 & & & & \\ & & 0 & 1 & & \\ & & -1 & 0 & & \\ & & & & & \\ 0 & & & & & \ddots \end{pmatrix} \tag{3.87}$$

That is,

$$M^T S M = S \tag{3.88}$$

The symplectic transformations therefore leave invariant the skew symmetric quadratic form $y^T S x$, where x and y are 2n dimensional real vectors.

There are also five simple groups, the exceptional groups, which do not belong to infinite sequences. They are denoted G_2, F_4, E_6, E_7, and E_8, where the subscript refers to the rank. The exceptional groups do not have any simple geometrical interpretation. The order (number of generators) of the simple groups are given in Table 3.1. Rules for constructing the structure constants may be found in [2.9]. Other useful properties are described in [3.33] and [3.34], for example.

Rank 4 Groups

The standard model gauge group $G_s = SU_3^c \times SU_2 \times U_1$ has rank 4. Therefore, any unified theory that contains G_s as a subgroup must have rank ≥ 4. (One can of course attempt partial unifications; for example $SU_3^c \times U_1$ [3.35, 3.36] or $SU_2 \times U_1$ [3.37] can be combined in groups of rank <4.) Georgi and Glashow [3.2] have classified all of the rank 4 groups that allow a single coupling constant. From Table 3.1 one can deduce that there are only nine of them, namely $[SU_2]^4$, $[SO_5]^2$, $[SU_3]^2$, $[G_2]^2$, SO_8, SO_9, Sp_8, F_4, and SU_5,

Cartan Label	Classical Group	Order (N)	Range of ℓ
A_ℓ	$SU_{\ell+1}$	$\ell(\ell+2)$	$\ell \geq 1$
B_ℓ	$SO_{2\ell+1}$	$\ell(2\ell+1)$	$\ell \geq 2$
C_ℓ	$Sp_{2\ell}$	$\ell(2\ell+1)$	$\ell \geq 3$
D_ℓ	$SO_{2\ell}$	$\ell(2\ell-1)$	$\ell \geq 4$
G_2	G_2	14	
F_4	F_4	52	
E_6	E_6	78	
E_7	E_7	133	
E_8	E_8	248	

Table 3.1 The simple Lie algebras. The subscript on the Cartan label is the rank (maximal number of simultaneously diagonalizable generators), and the order is the number of generators. Various classical groups are omitted because their algebras are equivalent to those in the table. These include $SO_6 \sim SU_4$, $SO_4 \sim SU_2 \times SU_2$, $SO_3 \sim SU_2$, $Sp_4 \sim SO_5$, and $Sp_2 \sim SU_2$.

where $[G]^2 \equiv G \times G$, etc. The first two are unacceptable because they do not contain an SU_3 subgroup. As will be discussed below, parity violation in the weak interactions requires either that the fermions be placed in a complex representation or that the number of fermions be doubled. If one requires the existence of complex representations then, of the nine candidates, only $[SU_3]^3$ and SU_5 are allowed.

For $[SU_3]^2$ one factor must be SU_3^c. But quarks, anti-quarks, and leptons have different color assignments, so they must belong to different representations of the weak SU_3 factor (which contains $SU_2 \times U_1$). However, one cannot pick a suitable electric charge operator Q unless additional quark and lepton fields are introduced (the sum of the quark charges must be zero since TrQ = 0, and the quarks and leptons must transform according to inequivalent representations in order to allow fractional quark charges and integer lepton charges). Furthermore, the criterion that there be a 1-1 correspondence of the representations of the factors is violated unless many exotic extra fields are introduced.

Therefore, the only acceptable rank 4 group that does not require additional fermions is SU_5. The Georgi-Glashow model is thus the "minimal" grand unified theory containing the standard model. Of course, one can consider larger theories (rank >4 or containing more fermions) or theories in which the standard model is modified.

Complex and Real Representations

From (2.3) it is apparent that if L^i form a representation of a group G then so do the conjugate representation matrices $-L^{i*}$. If L^i and $-L^{i*}$ are equivalent, i.e.,

$$-L^{i*} = U\, L^i\, U^\dagger,\ i = 1,\ \ldots,\ N, \tag{3.89}$$

where U is unitary, then the representation is said to be real (the term real is motivated by an alternate convention in which iL^i is the representation matrix). An example is the 2 and 2^* representations of SU_2 which are equivalent according to (2.139). If L^i and $-L^{i*}$ are not equivalent (such as the 3 and 3^* representations of SU_3) then the representation is said to be complex.

It is convenient in discussing grand unified theories to specify the representation matrices L^i_L of the left-handed fermions. Call this representation, which can be reducible, f_L. It is then apparent from (2.11) that the right-handed fields, which are related by $\psi_R = C\,\overline{\psi^c_L}^T$, transform according to $L^i_R = -L^{i*}_L$. That is $f_R = f^*_L$, where f^*_L is the conjugate representation to f_L. Therefore, if f_L is real the theory is vectorlike ($f_R \sim f_L$). If f_L is complex ($f_R = f^*_L \not\sim f_L$) the theory is chiral. QCD by itself is vectorlike, for example, because (for one flavor) f_L is the reducible representation

$$f_L = \underset{u_L}{3} + \underset{u^c_L}{3^*} \ , \tag{3.90}$$

while

$$f_R = f^*_L = \underset{u^c_R}{3^*} + \underset{u_R}{3} \sim f_L \ . \tag{3.91}$$

The standard model, on the other hand, is chiral. For one family

$$\begin{aligned} f_L = \ &\underset{\begin{pmatrix} u \\ d \end{pmatrix}_L}{(3,\ 2,\ \tfrac{1}{6})} + \underset{u^c_L}{(3^*,\ 1,\ -\tfrac{2}{3})} + \underset{d^c_L}{(3^*,\ 1,\ \tfrac{1}{3})} \\ &+ \underset{\begin{pmatrix} \nu \\ e^- \end{pmatrix}_L}{(1,\ 2,\ -\tfrac{1}{2})} + \underset{e^+_L}{(1,\ 1,\ 1)} \ , \end{aligned} \tag{3.92}$$

where the three entries represent the SU^c_3 and SU_2 representations and the weak hypercharge. Then

$$
\begin{aligned}
f_R = f_L^* = &\ (3^*, 2, -\tfrac{1}{6}) + (3, 1, \tfrac{2}{3}) + (3, 1, -\tfrac{1}{3}) \\
&\quad \begin{pmatrix} d^c \\ -u^c \end{pmatrix}_R \qquad\quad u_R \qquad\qquad d_R \\
&+ (1, 2, \tfrac{1}{2}) + (1, 1, 1) \\
&\quad \begin{pmatrix} e^+ \\ -\nu^c \end{pmatrix}_R \qquad e_R^- \\
&\neq f_L ,
\end{aligned}
\tag{3.93}
$$

where I have used equivalence of 2 and 2^*. In order to form a vectorlike generalization of the standard model (incorporating parity violation) it is necessary to introduce heavy fermions with opposite chiralities from the light fermions,

$$
\begin{pmatrix} U \\ D \end{pmatrix}_R \; U_R^c \; D_R^c \; \begin{pmatrix} N \\ E \end{pmatrix}_R \; N_R^c \; E_R^c \tag{3.94}
$$

(One can also add a (1, 1, 0) ν_L^c to f_L for symmetry.) These transform according to $f_{HL} = f_R$ and $f_{HR} = f_L$, so that the total representation of all the fermions is $f_L^{tot} = f_L + f_{HL} = f_L + f_L^*$, which is real.

When the standard model is embedded in a larger group G the left-handed fermion representation f_L can be either real or complex.

If f_L is real then the left-handed fermion representation of the standard model subgroup is automatically real also [3.38] so that parity violation in the weak interactions would require the introduction of heavy fermions as in (3.94). Georgi has argued against this possibility [3.38] on the grounds that $SU_3^c \times SU_2 \times U_1$ invariant mass terms, such as

$$-\mathscr{L}_m = m_1 (\overline{u}_L \overline{d}_L) \begin{pmatrix} U_R \\ D_R \end{pmatrix} + m_2 \overline{U}_L u_R + m_3 \overline{D}_L d_R + \dots \tag{3.95}$$

could be written in this case. He has speculated that the m_i (and hence the fermion mass eigenvalues) may be of the order of the grand unification scale $M_X \simeq 10^{14}$ GeV, which is clearly unacceptable. (This is an instance of the "survival hypothesis," to be discussed in Section 6.1.1.) This could come about if the m_i are generated by the same Higgs representations that break G down to G_S. For example, if right-handed 5* + 10 fields η_{Ra} and η_R^{ab} are added to the Georgi-Glashow model, then the quarks and leptons could acquire masses of order M_X from Yukawa couplings to adjoint Higgs representation

$$\begin{gathered} \overline{\psi}_L^a \quad \Phi_a^b \ \eta_{Rb} \\ \overline{\psi}_{Lab} \ \Phi_c^a \ \eta_R^{bc} . \end{gathered} \tag{3.96}$$

However, these terms would be eliminated if a discrete symmetry under $\eta_{Ra} \to -\eta_{Ra}$, $\eta_R^{ab} \to -\eta_R^{ab}$ were imposed on the Lagrangian. Bare mass terms of arbitrary magnitude could also be introduced unless forbidden by discrete symmetries.

Whether or not a reasonable fermion mass spectrum can be generated in vectorlike models, it is certainly true that the other alternative, namely that f_L is complex, is more economical in terms of fermion fields. Georgi and Glashow [3.2], Georgi [3.38], and Gell-Mann, Ramond, and Slansky [2.122, 3.25] have suggested the existence of complex representations as a criterion for grand unified theories. Gell-Mann et al. refer to such theories as flavor chiral. They discuss flavor chiral and vectorlike models and distinguish between models in which fermion number can or cannot be defined at the Lagrangian level.

The complex representations of the simple Lie algebras have been classified by Mehta and Srivastava [3.39]. The result is that only SU_n, $n > 2$, E_6, and SO_{4n+2} have complex representations.

Anomalies

The anomaly formula (2.209) was written for the convention of specifying the representations of the left- and right-handed fermions (under the assumption that fermions and anti-fermions belong to different representations). In the present convention of specifying the representation f_L of all left-handed fermions and anti-fermions (with matrices L_L), the formula becomes

$$A_{ijk} = 2A^L_{ijk} \equiv 2\mathrm{Tr}\ L^i_L\ \{L^j_L,\ L^k_L\} = -2A^R_{ijk} \tag{3.97}$$

where the last line follows from $f_R \sim f^*_L$. Georgi and Glashow [2.95] have defined "safe" representations as those for which $A^L_{ijk} = 0$. It is easy to verify from (3.97) that real representations (vectorlike theories) are safe.

Georgi and Glashow have shown that all representations of the orthogonal groups SO_n are safe except for $n = 6$ (SO_6 is not included in Table 3.1 because it has the same Lie algebra as SU_4). Gürsey, Ramond, and Sikivie [3.40] and Okubo [3.41] have shown that E_6 is safe. Hence, of all of the simple Lie groups only SU_n, $n > 2$, have unsafe representations. Okubo [3.41] and Banks and Georgi [3.42] have given formulas for the unsafe irreducible representations of SU_n.

Gauge theories based on SU_n, $n > 2$ must therefore either use safe representations or combine two or more unsafe irreducible representations which have anomalies that sum to zero. One way to do this is to combine a representation with its conjugate to form a real reducible representation (e.g.,

$f_L = 3 + 3^*$ in SU_3^c). A less trivial example is the Georgi-Glashow SU_5 model, in which the 5* and 10 anomalies are equal and opposite. This cancellation is not an accident: it is due to the fact that SU_5 is a subgroup of the safe group SO_{10}. The 16 of SO_{10} decomposes into 5* + 10 + 1 under the SU_5 subgroup. Since the anomalies for all SO_{10} generators, including the SU_5 subgroup, are zero, the 5* and 10 anomalies must cancel (the SU_5 singlet does not contribute). Other examples, in which the cancellation appears to be accidental, will be discussed in Section 6.1.2.

The Embedding of Color, Flavor, and Electric Charge

Gell-Mann, Ramond, and Slansky [3.25] have discussed the embedding of the color group SU_3^c in simple unified groups G. For each embedding of SU_3^c they define the flavor group G^{fl} as the maximal subgroup for which $G^{fl} \times SU_3^c \subset G$. G^{fl} presumably contains $SU_2 \times U_1$ as a subgroup. In particular they have considered the ansatz that all fermions belong to the 1, 3, or 3* representation of SU_3^c (i.e., that there are no "bizarre" fermions transforming as color sextets, octets, etc.). For each simple group they have described the possible SU_3^c embeddings and the associated flavor groups, as well as tabulating the representations that contain only 1, 3, and 3* of color.

They find that two of the exceptional groups fall outside of their assumptions: G_2 has rank two and G^{fl} is trivial. E_8 has no representations satisfying their ansatz. For the other groups, the embeddings fall into four classes. In the first two classes the quarks and leptons transform non-trivially with respect to different flavor groups (except for U_1 factors). In Class I, for example,

$$G^{fl} = G_\ell \times G_q \times U_1 \tag{3.98}$$

where the leptons are singlets w.r.t. G_q and the quarks are singlets w.r.t. G_ℓ. Weak universality is not a natural feature in such embeddings, because the quark and leptons couple to different gauge bosons. The $W^\pm$ must therefore be mixtures of the G_ℓ and G_q bosons and the observed universality of the quark and lepton weak couplings would have to somehow come about from the pattern of symmetry breaking.

Weak universality is much more natural in the other two classes. In Class III embeddings

$$G^{fl} = G_{q+\ell} \times U_1 \ , \tag{3.99}$$

where $G_{q+\ell}$ is a simple factor (or $SO_4 \sim SU_2 \times SU_2$) under which both quarks and leptons transform non-trivially. There are two cases of this type of embedding: (1) $G = SU_n$, with the fermions in the fundamental n or in the totally antisymmetric Kronecker product $[n, k] = (n^k)_A \equiv (n \times n \times \ldots \times n)_A$ of k n's. In this case, $G^{fl} = SU_{n-3} \times U_1$. (2) The other possibility is $G = SO_n$ with the fermions in a spinor representation. G^{fl} is then $SO_{n-6} \times U_1$.

Class IV embeddings have

$$G^{fl} = G_{\ell+q} \ , \tag{3.100}$$

with no U_1 factor to distinguish quarks and leptons. This occurs only for the exceptional groups F_4, E_6, and E_7, with the fermions in the lowest dimensional (26, 27, and 56) representations. $G_{\ell+q}$ is SU_3, $SU_3 \times SU_3$, and SU_6 for the three cases, respectively.

Gell-Mann, Ramond, and Slansky have also classified the possible choices for the electric charge operator Q in the various groups, under the assumption that quarks have fractional charges in the sequence (... 5/3, 2/3, -1/3, -4/3, ...) and that leptons have integer charges. For the first two classes of color embedding they find that there is considerable freedom in the choice of Q. The only restriction is that the sum of the charges of all fermions in a representation must be zero (since Tr $L_Q = 0$). For representations that are self-conjugate under antiparticle conjugation there are therefore no restrictions.

For Class III embeddings Q is strongly constrained. The quark charges, assumed to be in the sequence above, determine the lepton charges. Usually, the sum of the quark charges is non-zero. For the Georgi-Glashow model, for example, the only possible charge operator (that commutes with SU_3^c) is of the form

$$Q = \alpha\, T^3 + \beta Y \; ; \qquad (3.101)$$

α and β (and therefore the lepton charges) are uniquely determined by the requirement that $q_u = 2/3$ and $q_d = -1/3$. Also, $q_u + q_d \neq 0$.

For the Class IV embeddings (exceptional groups) the possible charge operators are severely limited. In particular, the sum of the quark charges must be zero, which would require a modification of the standard model quark structure. One could have two q = 2/3 quarks, u and c, and four charge -1/3 quarks, d, s, b, and h, for example (see Section 3.4.4). The quark charges specify the lepton charges in this class of theories. Color embeddings in the exceptional groups have also been discussed by Günaydin and Gürsey [3.43, 3.44].

C, P, and CP

Under C, P, and CP transformations, a left-handed fermion field ψ_L is changed to

$$\begin{aligned} C: &\quad \psi_L \to \psi_L^c \\ P: &\quad \psi_L \to \gamma_0 \psi_R \\ CP: &\quad \psi_L \to \gamma_0 \psi_R^c \end{aligned} \tag{3.102}$$

(Of course, ψ_L^c and ψ_R are related by $\psi_L^c = C\,\bar{\psi}_R^T$.) Hence, the left-handed fermion representation f_L is mapped onto itself under charge conjugation and onto $f_R = f_L^*$ under P and CP. (In writing (3.102) I have implicitly assumed that ψ_L^c and therefore ψ_R are defined. This must be the case for massive particles but need not be true for massless neutrinos. If ψ_L^c does not exist in the theory then C and P cannot be defined. CP is always defined, however, because ψ_L and $\psi_R^c = C\,\bar{\psi}_L^T$ are not independent fields.)

The gauge part of the Lagrangian is always CP invariant for a suitably defined CP transformation for the gauge fields $A_\mu^i \to A_\mu^{\prime i}$. For example,

$$\bar{\psi}_L\, \gamma^\mu\, L^i\, A_\mu^i\, \psi_L \underset{CP}{\to} \bar{\psi}_R^c\, \gamma_\mu\, L^i\, A_\mu^{\prime i}\, \psi_R^c = \bar{\psi}_L\, \gamma_\mu\, (-L^{iT}\, A_\mu^{\prime i})\, \psi_L \;, \tag{3.103}$$

so that the term is CP invariant for

$$-L^{iT}\, A_\mu^{\prime i} \equiv +L^i\, A^{i\mu} \;,$$

which also leaves $F_{\mu\nu}^i\, F^{i\mu\nu}$ invariant. Of course, CP violation can be introduced into the Yukawa couplings or Higgs potential, or it can be spontaneously broken.

The gauge terms need not be invariant under C and P, however, even when they are defined. In the standard model, for example, u_L, d_L, ν_L, and e_L^- are placed in doublets while u_L^c, d_L^c, ν_L^c (if it is introduced) and e_L^+ are in singlets, so that C and P are explicitly violated. In the Georgi-Glashow model u_L and u_L^c are arranged more symmetrically: they both appear in the same irreducible representation. However, d_L and e_L^+ are in a 10 while d_L^c and e_L^- are in a 5* so that C and P are again violated.

It is easy to modify the standard model (and the larger grand unified theory) so that the gauge couplings are C and P invariant. One can go to a vectorlike model, as in (3.94), for example. In this case the gauge couplings are invariant under

$$\begin{aligned} C:&\quad u_L \to U_L^c \\ &\quad A^{1,3} \to -A^{1,3}, \qquad A^2 \to +A^2 \\ P:&\quad u_L \to \gamma_0 U_R \\ &\quad A^{i\mu} \to A^i_\mu . \end{aligned} \tag{3.104}$$

These are perfectly good C and P transformations when viewed as transformations on the weak eigenstates. However, the fact that u_L is transformed into U_L^c and not into u_L^c indicates that C and P are violated explicitly or spontaneously by whatever mechanism generates the fermion masses.

Another possibility is to consider $SU_{2L} \times SU_{2R} \times U_1$ models in which $(u\, d)_L$ and $(u\ d)_R$ transform as (2, 1) and (1, 2), respectively, under the two SU_2 groups. In this case a C and P invariance in which the gauge bosons of the two SU_2 groups are transformed into each other can be imposed on the theory (so that the two gauge couplings are equal). C and P are broken by the mechanism that gives different masses to the bosons in the two SU_2 groups.

Gell-Mann and Slansky [3.45] have given an elegant general discussion of the cases in which invariant C and P transformations can be defined on the fermion and gauge fields. In SO_{10} and E_6 models, for example, the u_L, d_L, ν_L, and e^-_L fields appear in the same irreducible representation as u^c_L, d^c_L, ν^c_L, and e^+_L. It is then possible to define appropriate C and P transformations of the gauge fields so that the gauge terms are left invariant. These models contain the $SU_{2L} \times SU_{2R} \times U_1$ model as a subgroup, so that the SU_{2L} and SU_{2R} bosons are mapped into each other by the transformations. Slansky [2.45] has also discussed the charge conjugation properties of the weak isospin conserving part of the fermion mass matrix.

Fermion and Higgs Representations

One of the least attractive features of the SU_5 model is the highly reducible nature of the fermion representation, especially the repetition of 5* and 10 families. It would be very desirable to go to a theory in which all of the left-handed fermions are placed in a single irreducible representation (Georgi [3.38] has advocated the slightly less ambitious program of allowing the direct sum of several irreducible representations as long as no representation appears more than once). However, the large number of fermions in the standard model then requires that one must utilize a very large representation. One can either [2.122] use the fundamental or other low-lying representation of a very large group (i.e., include a family group in G), or one can use a higher-dimensional representation of a small group (such as SU_5). This possibility usually leads to the existence of fermions with "bizarre" quantum numbers. Another option which is suggested by but logically independent of the second possibility above is that the observed fermions (or at least the quarks) are themselves composites of "smaller" particles (e.g., of three fermions or a fermion and a boson) [2.172]. Some work along the first of these directions is described in Section 6.1.2.

The Higgs representations needed for SSB in large groups tend to be very large and the associated Higgs potentials are very complicated. For this reason the patterns of SSB have only been studied in relatively simple special cases (such as a single Higgs representation). The most extensive analysis is that of Li [3.6]. See also the discussions of specific models and the analysis of Michel and Radicati [3.46].

Fermion mass terms are of the form

$$\psi^T_{aL}\ C\ \psi_{bL} = \psi^T_{bL}\ C\ \psi^T_{aL} \quad . \tag{3.105}$$

Hence, if f_L is irreducible then only Higgs representations contained in the symmetric part of the direct product $f_L \times f_L$ can couple to fermion fields (and therefore generate masses) [2.122]. If f_L is the direct sum of identical irreducible representations the Yukawa couplings must be symmetric with respect to the simultaneous interchange of group and family indices.

3.4.2 SU_n Models

The simplest realistic grand unified theory is the Georgi-Glashow SU_5 model. As the strong and weak interaction subgroups of SU_5 are also associated with unitary groups, it is natural to consider the possibility of embedding the SU_5 model into still larger SU_n groups. Most of the extensions that have been considered have had the motivation of either enlarging the weak interaction subgroup or of adding a horizontal family group. The first class of models is considered below while the second is described in Section 6.1.2. The formalism of SU_n has been described in Section 3.2 and the patterns of SSB were considered in Section 3.3.1.

Enlarging the Weak Interaction Subgroup

A number of models have been proposed in which the $SU_{2L} \times U_1$ weak interaction subgroup is extended or modified. These include vectorlike SU_5 [3.47-3.49] and SU_6 [3.47-3.51] models, incorporating vectorlike $SU_2 \times U_1$ and $SU_3 \times U_1$ subgroups, respectively. Most of these models were motivated by considerations that are now obsolete (see Section 2.4.4) and are now ruled out by neutral current data. Viable variations could probably be constructed, but these are not motivated by any existing data.

Another problem with SU_6 broken down to $SU_3^c \times SU_3 \times U_1$ is that the gauge couplings of the two SU_3 factors will obey identical renormalization group equations [3.3] (to the extent that all the relevant masses are negligible). The observed difference between the strong and weak couplings would therefore suggest that: (a) $SU_3 \times U_1$ is broken to $SU_2 \times U_1$ at a very high mass scale (e.g., comparable to M_X); (b) that the pattern of fermion or Higgs masses distinguishes strongly between the two SU_3 groups; or (c) that the color group be extended, to SU_4^c for example, so that the strong coupling constant evolves more rapidly. SU_4^c could break down to SU_3^c at a mass scale large compared to present energies but small compared to M_X. Of course, the underlying gauge group would also have to be extended to accommodate SU_4^c.

One possible advantage of SU_n models is that versions can be constructed [3.47,3.49] in which the proton is stable (see Chapter 4). These must either be vectorlike or else must involve a large highly reducible fermion representation [3.47].

3.4.3 SO_n Models

The $n(n-1)/2$ generators of SO_n are $T^{ij} = -T^{ji}$, $i,j = 1, \ldots, n$, which satisfy [3.52-3.56]

$$[T^{ij}, T^{k\ell}] = i(-\delta^{jk} T^{i\ell} - \delta^{i\ell} T^{jk} + \delta^{ik} T^{j\ell} + \delta^{j\ell} T^{ik}) \tag{3.106}$$

from which one derives the field tensors

$$F^{ij}_{\mu\nu} = \partial_\mu A^{ij}_\nu - \partial_\nu A^{ij}_\mu + g(A^{ik}_\mu A^{kj}_\nu - A^{ik}_\nu A^{kj}_\mu) \ . \tag{3.107}$$

The defining representation of SO_n is the real n dimensional vector representation. The covariant derivative for a vector field Φ^i is

$$(D_\mu \Phi)^i = \partial_\mu \Phi^i - g A^{ij}_\mu \Phi^j \tag{3.108}$$

In addition to the n and the tensor representations derived from it (including the adjoint) SO_n has double valued representations called spinor representations. For SO_{2m-1} there is a single 2^{m-1} dimensional real spinor representation σ. For SO_{2m} there are two inequivalent spinors, each of dimension 2^{m-1}. If m is even the two spinors are real; while if m is odd the two spinors σ_+ and σ_- are complex and related to each other by conjugation (i.e., $\sigma_+ = \sigma_-^*$). The spinor representations are further discussed in [3.25, 3.52-3.56].

The covariant derivative for a spinor field χ^a, $a = 1, \ldots, 2^{m-1}$, is given (for $n = 2m$) by

$$(D_\mu \chi)^a = \partial_\mu \chi^a - \frac{ig}{4} \sigma^{ij}_{ab} A^{ij}_\mu \chi^b \ , \tag{3.109}$$

where the $2^{m-1} \times 2^{m-1}$ matrices σ^{ij}, $i,j = 1 \ldots, 2m$, are defined by

$$\sigma^{ij} = -\sigma^{ji} = \frac{1}{2i} [\gamma^i, \gamma^j] \ . \tag{3.110}$$

The generalized Dirac matrices γ^i are constructed in [3.52-3.56].

The SO_{10} Model

The SO_n groups are convenient for grand unification because they admit complex representations for $n = 2m$ (m odd) and are anomaly free (for $n \neq 6$). In particular, Georgi [3.57] and Fritzsch and Minkowski [3.58] have proposed a theory based on the rank-5 group SO_{10}, with each family of left-handed fermions assigned to a 16 dimensional complex spinor σ_+. In order to motivate the SO_{10} model it is convenient to consider two distinct subgroups, viz SU_5 and $SU_4 \times SU_2 \times SU_2$.

Under the SU_5 subgroup the fermion spinor σ_+ for the first family decomposes as

$$\begin{array}{ccccc} 16 \rightarrow 5^* & + & 10 & + & 1 \\ \begin{pmatrix} \nu_e & d^c \\ e^- & \end{pmatrix}_L & & \begin{pmatrix} e^+ & u & u^c \\ & d & \end{pmatrix}_L & & \nu^c_{eL} \end{array} , \tag{3.111}$$

so that each reducible 5* + 10 representation of the Georgi-Glashow model is combined in an IRREP of SO_{10} (this is why the anomalies of the 5* and 10 cancel). In addition σ_+ contains an SU_5 singlet field that has the quantum numbers of an antineutrino $\nu^c_L = C\ \bar{\nu}^T_R$. Hence, the SO_{10} model will in general involve massive neutrinos (see Chapter 6). Of the 45 generators of SO_{10}, 24 are those of the SU_5 subgroup. The other 21 generators connect or distinguish between the 5*, 10, and 1 in (3.111). In order to discuss their properties it is convenient to consider the subgroup decomposition

$$\begin{aligned} SO_{10} &\supset SO_6 \times SO_4 \\ &\sim SU_4 \times SU_2 \times SU_2 \\ &\supset SU^c_3 \times U'_1 \times SU_{2L} \times SU_{2R} , \end{aligned} \tag{3.112}$$

where $SU_4 \supset SU_3^c \times U_1'$ can be considered an extended color group with leptons as the fourth color [3.61] (see Section 2.5.2). With respect to $SU_3^c \times SU_{2L} \times SU_{2R}$, σ_+ decomposes as

$$\begin{array}{ccccccccc} 16 = & (3, 2, 1) & + & (1, 2, 1) & + & (3^*, 1, 2) & + & (1, 1, 2) \\ & \begin{pmatrix} u \\ d \end{pmatrix}_L & & \begin{pmatrix} \nu \\ e^- \end{pmatrix}_L & & \begin{pmatrix} d^c \\ -u^c \end{pmatrix}_L & & \begin{pmatrix} e^+ \\ -\nu^c \end{pmatrix}_L \end{array} \tag{3.113}$$

SO_{10} therefore contains the left-right symmetric $SU_{2L} \times SU_{2R} \times U_1'$ electroweak group (Section 2.4.4) as a subgroup. In particular, the right-handed fermions (left-handed antifermions) transform nontrivially under SU_{2R}.

The transformation properties of the 45 gauge bosons under $SU_3^c \times SU_{2L} \times SU_{2R}$ are

$$\begin{array}{ccccccccc} 45 = & (8, 1, 1) & + & (1, 3, 1) & + & (1, 1, 3) & + & (1, 1, 1) \\ & G_\beta^\alpha & & W_L^{\pm,0} & & W_R^{\pm,0} & & B' \\ & + (3^*, 2, 2) & + & (3, 2, 2) & + & (3, 1, 1) & + & (3^*, 1, 1) \\ & \begin{pmatrix} X & \overline{Y}' \\ Y & \overline{X}' \end{pmatrix} & & \begin{pmatrix} X' & \overline{Y} \\ Y' & \overline{X} \end{pmatrix} & & X_s & & \overline{X}_s \end{array} \tag{3.114}$$

The fifteen bosons G_β^α, $W_{L,R}^{\pm 0}$, and B' are associated with $SU_3^c \times SU_{2L} \times SU_{2R} \times U_1'$. The (unnormalized) U_1' generator is

$$Y' = 2(Y - T_R^3) \ , \tag{3.115}$$

where Y is the hypercharge operator of the $SU_{2L} \times U_1$ model. For fermions, Y' coincides with B-L, where B is baryon number and L is lepton number. Y' will sometimes be referred to as B-L. Therefore, the hypercharge generator Y of the GWS model is a linear combination of Y' and T_R^3. The other combination is the new diagonal generator associated with the extension of SU_5 to SO_{10}. X and Y form an SU_2 doublet (color antitriplet) of lepto-quark diquark bosons of electric charges 4/3 and 1/3, as in the SU_5 model. X' and Y' form a second doublet (color triplet) of bosons with charges 2/3 and -1/3; they can also contribute to nucleon decay. Finally, X_s is a color triplet lepto-quark boson with charge 2/3. It is associated with the extension of SU_3^c to SU_4^c. The X_s does not contribute to nucleon decay except by mixing with the X' [3.59], as shown in Figure 3.6. Such decays are of great interest because they violate B-L, but they are enormously suppressed by additional inverse powers of superheavy masses (Chapter 4).

The couplings of the X and Y to fermions are the same as in (3.39) and (3.86). The couplings of the X', Y', and X_s (for one family) are [3.59]

$$\begin{aligned}\mathscr{L} = \frac{g}{\sqrt{2}} X_\mu^{\prime\alpha} & [-\epsilon_{\alpha\beta\gamma} \bar{d}_L^{c\gamma} \gamma^\mu d_L^\beta - \bar{u}_{R\alpha} \gamma^\mu \nu_R^c - \bar{u}_{L\alpha} \gamma^\mu \nu_L^c] \\ + \frac{g}{\sqrt{2}} Y_\mu^{\prime\alpha} & [+\epsilon_{\alpha\beta\gamma} \bar{d}_L^{c\gamma} \gamma^\mu u_L^\beta - \bar{u}_{R\alpha} \gamma^\mu e_R^+ - \bar{d}_{L\alpha} \gamma^\mu \nu_L^c] \\ + \frac{g}{\sqrt{2}} X_s^\alpha & [\bar{d}_{L\alpha} \gamma^\mu e_L^- + \bar{d}_{R\alpha} \gamma^\mu e_R^- + \bar{u}_{L\alpha} \gamma^\mu \nu_L + \bar{u}_{R\alpha} \gamma^\mu \nu_R]\end{aligned} \tag{3.116}$$

It should be emphasized that the fermion fields in (3.116) are in the interaction basis. Mixing angles and phases would appear if (3.116) were re-expressed in the mass eigenstate basis. The diagrams for nucleon decay due to X' and Y' bosons are given by Machacek [3.59].

The Lagrangian of at least the fermion part of the SO_{10} model is C and P invariant. Under C the left-handed fermion representation σ_+ is mapped onto itself, and the bosons of the SU_{2L} and SU_{2R} subgroups are interchanged (see Section 3.4.1). Under P, σ_+ is mapped onto the conjugate representation $\sigma_- \sim \sigma_+^*$ of the right-handed fermions. C and P can be violated spontaneously (or explicitly in the Higgs sector of the model).

Patterns of Symmetry Breaking

SO_{10} has several interesting subgroups and symmetry breaking patterns [3.53,3.55-3.60,2.122]. In particular, it accommodates either the SU_5 Georgi-Glashow model or the $SU_4^c \times SU_{2L} \times SU_{2R}$ Pati-Salam model [3.61] as subgroups. The most detailed study is that of Rajpoot [3.56], who also considers the possibility of integer charged quarks.

Perhaps the simplest symmetry breaking pattern is [3.53,3.56]

$$SO_{10} \underset{M_G}{\rightarrow} SU_5 \underset{M}{\rightarrow} G_s \equiv SU_3 \times SU_2 \times U_1 \tag{3.117}$$

where M_G and M are typical masses associated with the two stages of symmetry breaking. (A third stage in which the $SU_3 \times SU_2 \times U_1$ is broken to $SU_3^c \times U_1^{EM}$ is implied.) The breaking of SO_{10} can be accomplished by either a spinor (16) Higgs respresentation or a 126 [2.122]. (An alternative [3.53,3.56] is to break SO_{10} to $SU_5 \times U_1$ via an adjoint (45) Higgs. The boson associated with the extra U_1 would affect neutral current processes if very light [3.62-3.63].) The second stage of breaking SU_5 to G_s can be carried out by an adjoint Higgs. If $M_G >> M$ then the phenomenology of nucleon decay, the prediction of $\sin^2\theta_W$, the existence of a desert, etc., are exactly as in the SU_5 model. Georgi and Nanopoulos [3.53] have emphasized the possibility that M_G and M may be comparable (in which case $SO_{10} \underset{M}{\rightarrow} G_s$ with $M \gtrsim 10^{14-15}$ GeV). Then, the spinor Higgs is characterized by one G_s singlet with VEV a, while the 45 has two G_s singlets (the analogues of combinations of the W_R^0 and B' in (3.114)), with VEV's b and c. Then [3.53,3.59],

$$M^2_{X,Y} \sim |b - c|^2$$

$$M^2_{X',Y'} \sim |a|^2 + |b + c|^2 \qquad (3.118)$$

$$M^2_{X_s} \sim |a|^2 + 4|b|^2 ,$$

so that if $|a|$, $|b|$, $|c|$, and $|b - c|$ are all comparable then X, Y, X', and Y' may all be important for nucleon decay.

Another interesting pattern is [3.56,3.60] the chain in (3.112). The first stage of symmetry breaking ($SO_{10} \rightarrow SU_4 \times SU_2 \times SU_2$) can be implemented by a 120 or 54 of Higgs. It is characterized by a mass scale M_G of 10^{14-15} GeV. Nucleon decay is associated with bosons with this mass scale, so the proton lifetime is similar to the SU_5 model. The breaking to $G' \equiv SU_3^c \times SU_{2L} \times SU_{2R} \times U_1'$, due to an adjoint of Higgs, is characterized by a mass M_c. Rajpoot then considers two possibilities:

$$\text{i)} \quad G' \underset{M \sim M_W}{\rightarrow} G_s$$

$$\text{ii)} \quad G' \underset{M_R}{\rightarrow} SU_3^c \times SU_{2L} \times U_1^R \times U_1' \underset{M}{\rightarrow} G_s \qquad (3.119)$$

Chain i) does not lead to small enough values of $\sin^2\theta_W$ for any values of M_G and M_c below the Planck mass. In Chain ii), U_1^R is generated by T_R^3. Rajpoot finds reasonable values for $\sin^2\theta_W$ and α_s for $M_c \sim 10^{12}$ GeV and $M_R \sim 10^9$ GeV (for $M \gtrsim M_W$). The final breaking down to G_s can be due to a Higgs spinor. If M is very close to M_W the extra U_1 factor will modify the neutral current of the GWS model, as has been discussed by Deshpande and Iskandar [3.63] and Masiero [3.62]. A pattern similar to ii), but without the extra U_1 factor, has been considered by Shafi, Sondermann, and Wetterich [3.64]. They find results similar to Rajpoot and discuss the phenomenological constraints on the

X_s boson mass ($\simeq M_c$). Still another variation on this chain is discussed by Lazarides et al. [3.65].

A third symmetry breaking pattern is [3.53,3.56]

$$\begin{aligned} SO_{10} &\underset{M_G}{\rightarrow} SU_4^c \times SU_{2L} \times U_1^R \\ &\underset{M_c}{\rightarrow} SU_3^c \times SU_{2L} \times U_1^R \times U_1' \\ &\underset{M}{\rightarrow} G_s , \end{aligned} \tag{3.120}$$

with the first two stages due to adjoint Higgs and the third to a spinor. Georgi and Nanopoulos [3.53] and Rajpoot [3.56] find that values of $\sin^2\theta_W \sim 0.23$ (larger than the SU_5 prediction) can be accommodated for $M_G \gtrsim 10^{14\text{-}15}$ GeV, $M_c \gtrsim 10^{12}$ GeV, and $M \gtrsim M_W$ (the effects on neutral currents are similar to the previous case).

We see that the possible symmetry breaking patterns in SO_{10} and larger groups are very complicated and depend on the details of the Higgs representations. A detailed study of the minimization of the Higgs potential for an SO_{10} model with a 45 + 16 + 16* Higgs representation has been given by Buccella et al. [3.66].

Fermion Masses

With the left-handed fermions assigned to spinor representations ψ_L, Yukawa couplings must be of the form

$$\overline{\psi}_R^c \, \psi_L \, \varphi = \psi_L^T \, C \, \psi_L \, \varphi \ , \tag{3.121}$$

so that $\varphi^\dagger$ must belong to [3.53,2.122,3.67]

$$16 \times 16 = 10 + 126 + 120 \ . \tag{3.122}$$

The couplings of the neutral, color singlet Higgs components (which can have nonzero VEV's) have the form [3.67]

$$
\begin{aligned}
\overline{\psi}_R^c\, \psi_L\, \varphi_{10} &= \varphi_{10}(5)(\overline{u}_R u_L + \overline{\nu}_R \nu_L) \\
&\quad + \varphi_{10}(5^*)(\overline{d}_R d_L + \overline{e}_R e_L) \\
\overline{\psi}_R^c\, \psi_L\, \varphi_{126} &= \varphi_{126}(1)\ \nu_R\, C\, \nu_R + \varphi_{126}(15)\ \nu_L\, C\, \nu_L \\
&\quad + \varphi_{126}(5)(\overline{u}_R u_L - 3\overline{\nu}_R \nu_L) \\
&\quad + \varphi_{126}(45^*)(\overline{d}_R d_L - 3\overline{e}_R e_L) \\
\overline{\psi}_R^c\, \psi_L\, \varphi_{120} &= \varphi_{120}(5)\ \overline{\nu}_R \nu_L + \varphi_{120}(45)\ \overline{u}_R u_L \\
&\quad + \varphi_{120}(5^*)(\overline{d}_R d_L + \overline{e}_R e_L) \\
&\quad + \varphi_{120}(45^*)(\overline{d}_R d_L - 3\overline{e}_R e_L)\ .
\end{aligned}
\tag{3.123}
$$

In (3.123) the SU_5 content of the relevant components of the Higgs fields are listed in parentheses. The φ_{10} and φ_{126} couplings are symmetric in the family indices (i.e., they generate symmetric mass matrices), while the φ_{120} couplings are antisymmetric.

The couplings of φ_{10} generate equal d and e mass matrices, just as in the SU_5 model with 5's of Higgs. The degeneracy can be broken by incorporating a φ_{126} or φ_{120} with the appropriate (45) components nonzero. The situation for these mass matrices is therefore similar to the SU_5 model, and will be discussed further in Chapter 6. (Incidentally, the symmetry breaking pattern in (3.120) can preserve the successful prediction of m_b/m_τ [2.53].) One of the motivations for a detailed consideration of the SO_{10} model was the possibility of imposing additional symmetries to restrict the Yukawa couplings, so that M^u would be related to M^d. Georgi and Nanopoulos [2.53] constructed a model in which a global U_1^3 symmetry (softly broken to avoid unwanted Goldstone bosons) was imposed. The t quark mass was predicted ($\simeq$ 14 GeV), along with various mixing angles, but the relation appears to be unsuccessful.

φ_{10} yields equal Dirac mass matrices for the u quarks and neutrinos, which by itself would be a disaster. The problem of neutrino masses will be discussed more fully in Chapter 6, but ways to achieve small masses include: (a) Introduce [3.53] an SO_{10} singlet field E_L. E_L and ν_R can be given a very large Dirac mass via a spinor Higgs, leaving one massless Majorana particle (basically ν_L). (b) Let $\varphi_{126}(1)$ be very large (of order of M_X) [2.122,3.68] so that ν_R develops a superheavy Majorana mass. ν_L will remain essentially massless. The $\varphi_{126}(1)$ will of course break SO_{10} down to SU_5. (c) Utilize 120 or 126 dimensional Higgs fields [3.69].

Other SO_n Models and Conclusion

Larger SO_n models which combine SO_{10} with horizontal and/or technicolor interactions will be discussed in Chapter 6. Vectorlike SO_{11} [3.70] and SO_{12} [3.71] models have also been constructed.

The SO_{10} model is an attractive extension of the SU_5 model in that the u, d, e, and ν fermions are treated much more symmetrically. Also, there is more freedom in choosing symmetry breaking patterns, so that a larger value of $\sin^2\theta_W$ and new thresholds below 10^{14} GeV are possible. Chanowitz et al. [3.60] have argued that SO_{10} is the only attractive rank 5 extension of SU_5, because SU_2^5 does not contain SU_3^c, while SO_{11}, SU_6, and Sp_{10} do not have suitable representations to combine 5* + 10 without introducing additional fermions.

3.4.4 The Exceptional Groups [3.72]

The groups SO_n, SU_n, and Sp_{2n} are members of infinite sequences. There is no a priori reason to expect SU_5 to be a better candidate for a grand unified theory than SU_6, SU_7, ... The exceptional groups, on the other hand, have the esthetic advantage of uniqueness -- there are only five of them, G_2, F_4, E_6, E_7, and E_8.

Of the exceptional groups, G_2 is not large enough to contain the standard model as a subgroup. Only E_6 admits complex representations, but if one also considers vectorlike theories then F_4, E_6, E_7, and E_8 are all candidates for grand unification.

E_8 has the interesting feature that the lowest dimensional representation is the adjoint (E_8 is the only simple Lie group with this property). However, the adjoint is 248 dimensional, so an enormous number of gauge bosons and fermions are required. There are no representations that have color singlets, triplets, and anti-triplets only [3.25]. For example, if the fermions are placed in the 248 then there will be a one-to-one correspondence between the gauge bosons and the fermions, implying a charge -4/3 quark, a color octet of quarks, etc. In fact, in the absence of Higgs particles there is a supersymmetry between the fermions (in a 248) and the gauge bosons [3.73-3.74] in E_8. A number of E_8 models [3.72-3.74] have been discussed, which differ in their embeddings of $SU_2 \times U_1$ (and therefore in their predictions for $\sin^2\theta_W$, the fermion charges, etc.). In particular, symmetry breaking patterns that start with the maximal subgroups $E_6 \times SU_3$ [3.72], $E_7 \times SU_2$ [3.73], SU_9 [3.73a], and SO_{16} [3.74] have been considered. For example, Bars and Günaydin [3.74] describe the chain $E_8 \to SO_{16} \to SO_{10} \times SO_6 \to SU_5 \times SU_3$, with the fermions and

Higgs fields in a 248 and 3875, respectively. A specific assumption is made concerning the relative magnitudes of the VEV's of the various components of the 3875. This leads to the prediction of three light SU_5 families (5* + 10) and three somewhat heavier (<1 TeV) conjugate families (5 + 10*). All other fermions are superheavy. $\sin^2\theta_W$ is 3/8 at the SU_5 unification mass.

The F_4, E_6, and E_7 groups each have a single representation (always the lowest dimensional, or "fundamental") [3.25] involving 1, 3, and 3*'s of color only (i.e., no bizarre fermions). They are all desirable for unification because weak universality emerges naturally (i.e., the quarks and leptons both transform nontrivially under the same weak subgroup, with no U_1 factor to distinguish them; see Section 3.4.1). Furthermore, the electric charge operators are essentially uniquely determined (Section 3.4.1), with the sum of the quark charges being zero and the quark charges determining the lepton charges [3.25].

However, the maximal subgroup of F_4 is $SU_3^c \times SU_3$, under which the fundamental 26 decomposes as

$$26 \rightarrow (1,\ 8) + (3,\ 3) + (3^*,\ 3^*) \ . \tag{3.124}$$

Therefore F_4 generates a vectorlike theory with the quarks and leptons in different representations of the weak SU_3 group, leading to a proliferation of undesired quarks and leptons. Furthermore, $\sin^2\theta_W$ is predicted [3.72] to take the unacceptably large value of 3/4 at the unification mass.

The E_7 model [3.75-3.77] has the advantage that all known fermions can be assigned to a single fundamental 56 dimensional representation. In the simplest color embedding [3.75,3.25]

$$E_7 \rightarrow SU_3^c \times SU_6 \ , \tag{3.125}$$

with

$$56 \to (1,\ 20) + (3,\ 6) + (3^*,\ 6^*)$$
$$133 \to (1,\ 35) + (8,\ 1) + (3,\ 15^*) + (3^*,\ 15)\ , \qquad (3.126)$$

where the 133 is the adjoint. There are two charge 2/3 quarks (u and c) and four charge -1/3 quarks (d, s, b, h), with no t quark. There are four charge -1 leptons and 12 two-component neutral leptons. Unfortunately, the simplest E_7 model, with a single 56, fails [3.78] when confronted with the charged and neutral current data. Among the problems are: (a) $\sin^2\theta_W$ is predicted to be [3.75] 3/4, renormalized down to $\simeq 2/3$ if E_7 is broken down to $SU_3^c \times SU_2 \times U_1$ at superheavy masses [3.79,3.75] or to $\simeq 1/2$ if the breaking is to $SU_3^c \times SU_2 \times SU_2 \times U_1$ [3.75]. Both values are much too large to be acceptable. (b) The model is vectorlike, with u_R in a weak doublet with either b_R or h_R or a mixture, implying anomalies in charged current $\bar{\nu}$ scattering and an unacceptable neutral current [2.70]. (c) There is a danger of FCNC in the charge -1/3 sector. (d) The leptonic sector is complicated, with two left-handed SU_2 doublets, two right-handed doublets, and left- and right-handed triplets. There are several ways to assign the leptons [3.75-3.76], but it is difficult to accommodate the correct leptonic neutral current and the V-A decay of the τ.

Some of these objections (those involving the charged current) could possibly be overcome by adding additional 56's and many heavy particles to fill them, but then one of the main advantages of E_7 (a single fermion representation) would be lost. Lednický and Tseitlin [3.80] have suggested a variant on the model in which the low energy weak group is $SU_2 \times U_1^3$. It may then be possible to fit the neutral current data by adjusting six parameters, but the success of the GWS model would then be an accident.

The E_6 Group

It therefore appears that E_6 is the most attractive exceptional group for grand unification. Only the fundamental 27 (and its conjugate) satisfy the constraint of having only 1, 3, and 3* representations of the color subgroup. Under the maximal subgroup $SU_3^c \times SU_3 \times SU_3$, the 27 and the adjoint (78) transform as

$$\begin{aligned} 27 &\to (1, 3^*, 3) + (3, 3, 1) + (3^*, 1, 3^*) \\ 78 &\to (8, 1, 1) + (1, 8, 1) + (1, 1, 8) \\ &\quad + (3, 3^*, 3^*) + (3^*, 3, 3) \ . \end{aligned} \tag{3.127}$$

Another interesting subgroup is $SO_{10} \times U_1$. (The chain $E_6 \supset SO_{10} \supset SU_5$ is especially interesting because SU_5 and SO_{10} can be considered to be the E_4 and E_5 members of the E_n sequence [3.34].) Under SO_{10},

$$\begin{aligned} 27 &\to 16 + 1 + 10 \\ 78 &\to 45 + 16 + 16^* + 1 \ . \end{aligned} \tag{3.128}$$

Most E_6 models utilize the complex 27 dimensional representation for the left-handed fermions. (A vectorlike E_6 model utilizing a 27 + 27* shares most of the undesirable features of the E_7 model, into which it can be embedded [3.81].) The models can be divided into two classes, depending on whether they emphasize the $SU_3^c \times SU_3 \times SU_3$ or the SO_{10} subgroups.

Models in the first category [3.81-3.87] are more economical in the number of fermions, with all left-handed fermions assigned to two 27's. Early versions of the model by Gürsey, Ramond, and Sikivie [3.82] and Gürsey and Serdaroğlu [3.83] possessed purely vector neutral current couplings for the electron and muon, but this was modified in later models by Gürsey [3.83], Achiman and Stech [3.84], and Shafi [3.85]. These models contain two charge

2/3 quarks (u, c), four charge -1/3 quarks (d, s, b, h), four charge -1 leptons (e^-, μ^-, τ^-, M^-), and ten two-component neutrinos (ν_L^e, ν_L^μ, ν_L^τ, ν_L^M, $N_L^{c\tau}$, N_L^{cM}, α_L^{ce}, $\alpha_L^{c\mu}$, β_L^e, β_L^μ). In the Gürsey model, the two 27's are:

$$(1, 3^*, 3) \quad + (3, 3, 1) + (3^*, 1, 3^*)$$

$$\begin{pmatrix} N^c_{\tau L} & \tau_L^+ & e_L^+ \\ \tau_L^- & \nu_{\tau L} & \alpha^c_{eL} \\ e_L^- & \nu_{eL} & \beta_{eL} \end{pmatrix} \quad (u_L \ d_L \ b_L) \qquad \begin{pmatrix} u_L^c \\ d_L^c \\ b_L^c \end{pmatrix} \tag{3.129}$$

$$\begin{pmatrix} N^c_{ML} & M_L^+ & \mu_L^+ \\ M_L^- & \nu_{ML} & \alpha^c_{\mu L} \\ \mu_L^- & \nu_{\mu L} & \beta_{\mu L} \end{pmatrix} \quad (c_L \ s_L \ h_L) \qquad \begin{pmatrix} c_L^c \\ s_L^c \\ h_L^c \end{pmatrix},$$

where the first SU_3 factor of the weak subgroup acts horizontally and the second vertically. Hence, the weak group can be identified as $SU_{3L} \times SU_{3R}$, where SU_{3L} acts nontrivially on the left-handed quarks and SU_{3R} acts nontrivially on the right-handed quarks (left-handed antiquarks). The leptons transform nontrivially under both. The $SU_{3L} \times SU_{3R}$ weak group has been discussed by Achiman [3.86] and Minkowski [3.87]. Under the $SU_{2L} \times U_1$ subgroup, the doublets are (up to mixing effects)

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \qquad \begin{pmatrix} c \\ s \end{pmatrix}_L$$

$$\begin{pmatrix} \nu_{eL} \\ e_L^- \end{pmatrix} \begin{pmatrix} \nu_{\mu L} \\ \mu_L^- \end{pmatrix} \begin{pmatrix} \nu_{\tau L} \\ \tau_L^- \end{pmatrix} \begin{pmatrix} N^c_{\tau L} \\ \tau_L^+ \end{pmatrix} \begin{pmatrix} \nu_{ML} \\ M_L^- \end{pmatrix} \begin{pmatrix} N^c_{ML} \\ M_L^- \end{pmatrix} \tag{3.130}$$

The major predictions of the model are: (a) There is no t quark. (b) There is a fourth charge -1/3 quark. (c) There is a fourth charged lepton. (d) The τ^- (and M^-) neutral currents are pure vector. (e) In addition, the ordinary neutrinos will generally be massive [3.34,3.81-3.85, 3.88] and there will be numerous additional neutral leptons (which may be very massive). (f) Mixing between the d, b, h, and s and between the e, μ, τ, and M will generally lead to FCNC. If the mixing of the b and h with the d and s is somehow suppressed (to avoid FCNC) then the b and h would not be able to decay via the ordinary $SU_2 \times U_1$ charged current weak interactions.

It has been suggested [3.81,3.83-3.84] that the dominant symmetry breaking mechanism may be due to an adjoint Higgs representation in which the (1, 8, 1) and (1, 1, 8) components have VEV's along the hypercharge (8) directions of the two SU_3 groups. In this case, $E_6 \to SU_3^c \times SU_{2L} \times U_{1L} \times SU_{2R} \times U_{1R}$. (However, Harvey has recently argued [3.89] that the minimum of the Higgs potential for a single adjoint representation would, when radiative corrections are considered, be in a direction to break E_6 to $SO_{10} \times U_1$ instead.) Components of the same or additional adjoints in the 3 and 6 directions of the SU_3 subgroups, or of other Higgs representations, could break the symmetry further, to $G_s \equiv SU_3^c \times SU_2 \times U_1$, for example. In any case, it is assumed that the bosons which mediate nucleon decay acquire superheavy masses so that the predictions for nucleon decay are similar to those of the SU_5 and SO_{10} models. If E_6 is broken to G_s at superhigh energies, then the predictions for $\sin^2\theta_W$ and α/α_s depend only on the G_s content of the fields. For the fields in (3.129), α/α_s and

$$\sin^2\theta_W = \frac{Tr(T^3)^2}{Tr(Q)^2} \tag{3.131}$$

are both predicted to be 3/8 at $Q^2 \simeq M^2$, just as in the SU_5 model. The renormalization effects are also the same (at least to leading order and ignoring the Higgs contributions) so that $\sin^2\theta_W \simeq 0.20$ at low energies. Achiman and Stech [3.84] and Shafi [3.85] have argued that other patterns in which E_6 is broken to a group larger than G_s at M (such as $SU_3^c \times SU_{2L} \times U_{1L} \times SU_{2R} \times U_{1R}$ or $SU_3^c \times SU_{2L} \times SU_{2R} \times U_1'$) and then to G_s at low energies will lead to larger values of $\sin^2\theta_W$ in the range 0.25-0.30 (see Section 6.2).

Higgs fields that can contribute to fermion masses must be contained in

$$27 \times 27 = 27^*_S + 351_S + 351'_A , \tag{3.132}$$

where S and A refer to symmetric and anti-symmetric (the 351 and 351' are of course inequivalent). There are many electrically neutral color singlet components in these representations (the 27 contains five such components, while the 351_S contains nineteen, with five transforming as (1, 3*, 3) and fourteen as (1, 6*, 6) under the $SU_3^c \times SU_3 \times SU_3$ subgroup), so in general there are few predictions for the fermion masses. Gürsey [3.83] in particular has shown that there are combinations of VEV's which lead to a reasonable fermion mass spectrum. The existence of a 351 appears necessary [3.83-3.84] and it is extremely difficult to keep the ordinary neutrinos massless [3.84,3.88].

Another class of E_6 models [3.84,3.90-3.92] focusses on the $SO_{10} \times U_1$ subgroup. It incorporates the canonical $SU_{2L} \times U_1$ assignments for the light fermions at the expense of introducing additional heavy fermions. Each 27 of fermions decomposes under SO_{10} into 16 + 1 + 10. The 16 is the standard SO_{10} representation for a single fermion family, while the SO_{10} singlet is a neutral lepton E_L. The E_L can combine with the SU_5 singlet ν_L^c to form a massive Dirac particle, as suggested by Georgi and Nanopoulos [3.53]. (They could also have large Majorana masses). The fermions in the 10 are assumed to somehow acquire a large mass. At least three 27's are required to incorporate the known fermion families. Most of the physics in these models is similar to that of the SO_{10} model.

Barbieri and Nanopoulos [3.92] have recently given a very attractive model in which E_6 is broken down to $G_s = SU_3^c \times SU_2 \times U_1$ by 351_S and $351_A'$ Higgs representations. They argue that the unwanted fermions (the 10, the E_L, and the ν_L^c) form a vectorlike representation w.r.t. the G_s subgroup. They therefore can and in general will receive superheavy G_s invariant masses by the same Higgs fields that break E_6. (This is the survival hypothesis -- see Chapter 6.) The other 15 fermions in each 27 can only receive masses from the 27 dimensional Higgs which breaks G_s to $SU_3^c \times U_1^{EM}$ and which therefore is presumed to have a very small VEV. The desired pattern of fermion masses is therefore tied to the (still unexplained) hierarchy of gauge boson masses.

The mass matrix generated by the 27 for the light fermions would by itself imply $M^e = M^d$, just as in the SU_5 and SO_{10} models. However, in this case there are large off-diagonal couplings between the light and heavy fermions (generated by the 351 and 351') which can modify the prediction $m_d/m_s = m_e/m_\mu$. Unfortunately, the desirable relation between m_b and m_τ will also be lost except for special limits of the parameters.

Ramond [3.90] and Barbieri and Nanopoulos [3.92] have noted that the necessary symmetry breaking in the E_6 model can be implemented by the Higgs representations (27*, 351, 351') contained in the direct product 27 × 27 (e.g., no adjoint is needed). This feature, which is not present in the SU_5 or SO_{10} models, is interesting because it suggests that the symmetry breaking could be dynamical, with the Higgs fields replaced by fermion-antifermion bound states (see Sections 2.5.3 and 6.8). For example, the fields transforming as 16 × 1 under the SO_{10} subgroup can break E_6 down to SU_5; the fields transforming as 10 × 10 contain an adjoint of SU_5 (in the 351 or 351' of E_6) which can break SU_5 to G_s.

Stech [3.72] has recently discussed fermion masses and symmetry breaking patterns in the E_6 model.

3.4.5 Semi-Simple Groups

Many models have been based on semi-simple groups. Typical features of these models are: (a) baryon number B, lepton number L, and fermion number F = 3B + L are usually exact global or local symmetries of the Lagrangian. They are violated spontaneously if at all. (b) Parity violation is usually spontaneous. (c) The unification mass M is often very low (e.g., $M \sim 10^{4-6}$ GeV), so that no large desert or extreme hierarchy exists. (d) If B, L, and F are conserved in the Lagrangian, the proton will either be stable or will decay via mixing effects induced by SSB. In the latter case, the suppression will hopefully be sufficient to yield a long lifetime despite the low unification mass.

Maximal Gauge Groups

Consider a theory in which n left-handed fermions and n left-handed anti-fermions (right-handed fermions) are distinguished in the Lagrangian by a conserved fermion number F. If F is a global quantum number then the maximal possible gauge symmetry of the fermions is [3.93-3.94] $SU_{nL} \times SU_{nR}$, where SU_{nL} acts on the n fermions and SU_{nR} acts on the n antifermions. Parity is conserved, at least in the fermion sector of the Lagrangian if the two factors have the same gauge coupling. If F is gauged then the maximal symmetry is SU_{2n}, with n fermions and n antifermions combined in the fundamental representation. F can be violated spontaneously in this case.

For example, for two families one has n = 16 (n^α, d^α, c^α, s^α, ν_e, e^-, ν_μ, μ^-) so that the maximal group is SU_{32} or $SU_{16L} \times SU_{16R}$, which has been studied by Fritzsch and Minkowski [3.93]. For one family (n^α, d^α, ν_e, e^-) one has SU_{16} or $SU_{8L} \times SU_{8R}$ [3.93].

One problem with these maximal symmetry groups is that they contain anomalies. It is necessary to add additional "mirror" representations of heavy fermions that transform according to the conjugate representation (i.e., with the opposite chirality) to cancel the anomalies.

A recent $SU_{8L} \times SU_{8R}$ model by Pirogov [3.95] illustrates the basic ideas: the left- and right-handed fermions $(u^\alpha, d^\alpha, \nu_e, \mu^-)_{L,R}$ transform as (8, 1) and (1, 8), respectively, with additional families assigned to similar representations. The model as written has anomalies, so heavy mirror families of fermions in which the L and R fields transform as (1, 8) and (8, 1), respectively, must be introduced to make the theory vectorlike. (Of course, the theory is not really maximally gauged because of the repetition of light and mirror families.) At a unification mass M, G is broken to $G^S \times G^W$, where G^S and G^W are the strong and weak subgroups. There are many breaking patterns. For example, G^S could be SU_3^c or $SU_{3L}^c \times SU_{3R}^c$. In the latter case one has chiral color [3.96], in which there are separate gluons coupling to left- and right-handed quarks. $SU_{3L}^c \times SU_{3R}^c$ would subsequently break down to SU_3^c, with the axial gluons acquiring a mass of perhaps 1 GeV. Similarly, one could have $G^W = SU_{2L} \times SU_{2R} \times U_{1L} \times U_{1R}$,

$SU_{2L} \times U_1$, etc. For the case $G^S = SU^c_{3L} \times SU^c_{3R}$, $G^W = SU_{2L} \times SU_{2R} \times U_{1L} \times U_{1R}$, Pirogov obtains a unification mass of $\simeq 10^{6-8}$ GeV with $\sin^2\theta_W \simeq 1/3$ at low energies. The proton is apparently stable for the fractionally charged quark version of this model.

This model should not be taken too seriously (because of the large value for $\sin^2\theta_W$), but it illustrates the possibility of having a small unification mass if G^S and G^W are sufficiently complicated and the Lagrangian conserves B and L.

Another $SU_{8L} \times SU_{8R}$ model [3.97] will be discussed in connection with CP violation in Chapter 6, as will semi-simple groups proposed in connection with horizontal symmetries. A non-gauge phenomenological group based on $SO_8 \times SO_8$ has also been studied [3.98]. All fermions and antifermions are united in a single real (8, 8) representation.

The Pati-Salam Models

Pati and Salam [3.94,3.99-3.100] have emphasized a series of models of the form $G = G^S \times G^W$, where G^S and G^W are identical strong and weak groups related by a discrete symmetry. These models have been thoroughly reviewed by Pati [3.94], so I will only outline some of the basic ideas here.

The simplest model is based on the group $G = SU_4^4$. The 16 left-handed and 16 right-handed fermions of the first two families (including right-handed neutrinos) are assigned to the representations

$$f_{L,R} = \begin{pmatrix} u^R & u^G & u^B & \nu_e \\ d^R & d^G & d^B & e^- \\ s^R & s^G & s^B & \mu^- \\ c^R & c^G & c^B & \nu_\mu \end{pmatrix}_{L,R} , \qquad (3.133)$$

with $f_{L,R}$ transforming as $f_L \sim (4, 1, 4^*, 1)$ and $f_R \sim (1, 4, 1, 4^*)$ under the four SU_4 factors. G can be interpreted as $G^S \times G^W$, where $G^S = SU_{4L} \times SU_{4R}$ is a chiral color group (with separate gluons coupling to L and R fermions) [3.96] and with lepton number as the fourth color [3.99]. G^S acts horizontally on (3.133). Similarly, $G^W = SU_{4L} \times SU_{4R}$ is a left-right symmetric chiral weak group which acts vertically (Rajpoot [3.101] has recently given an SU_4^2 version of the model with $G^S = G^W = SU_4$).

Additional representations of heavy mirror fermions are required to cancel anomalies. These transform like f_L and f_R, but with the helicity labels reversed, so the theory is vectorlike. Depending on the detailed form of the discrete symmetry used to relate the four factors (to ensure a single coupling constant) there may or may not be additional representations [3.102] transforming as $(4, 1, 1, 4^*)$, $(1, 4, 4^*, 1)$, $(4, 4^*, 1, 1)$, and $(1, 1, 4, 4^*)$. Additional families of fermions: (a) may be incorporated in these extra representations or the mirror representations; or (b) may be included by repeating the representations in (3.133); or (c) may be incorporated in $f_{L,R}$ by extending the group to SU_6^4, SU_8^4, etc. [3.103].

$G = SU_4^4$ is a chiral group as far as the light fermions are concerned (although it is vectorlike when the mirror fermions are included). Parity is violated spontaneously. Fermion number F is associated with a global symmetry in SU_4^4, since the left-handed fermions and antifermions appear in different representations. Of course, SU_4^4 can be extended to larger groups in which the fermions and antifermions are combined in a single representation, such as the maximal group SU_{32} [3.94]. B and L are linear combinations of F and a local symmetry generator. B and L may or may not be spontaneously violated, depending on the pattern of symmetry breaking. F is not spontaneously violated in the SU_4^4 theory, but can be in extensions of the theory.

The Unification Mass

The SU_4^4 model possesses the advantages listed at the start of this section. In particular, one possible symmetry breaking chain is

$$\begin{aligned} G &= \left(SU_{4L}^c \times SU_{4R}^c\right) \times \left(SU_{4L} \times SU_{4R}\right) \\ &\xrightarrow[M]{} \left(SU_{3L}^c \times SU_{3R}^c \times U_{1L} \times U_{1R}\right) \times \left(SU_{2L} \times SU_{2R}\right) , \end{aligned} \tag{3.134}$$

where $SU_{3L}^c \times SU_{3R}^c$ is a chiral color group and $SU_{2L} \times SU_{2R}$, when combined with the U_1' factor associated with the vector generator in $U_{1L} \times U_{1R}$, is a left-right symmetric electroweak group. $SU_{2L} \times SU_{2R} \times U_1'$ can then break down to $SU_{2L} \times U_1$ at $M_{W_R} \gtrsim 5\ M_W$, implying the spontaneous violation of parity. The chiral color group breaks down to SU_3^c at a mass $M_A < M_W$. The mass of the axial vector gluon (M_A) is often taken to be $\simeq$1 GeV.

Elias, Pati, and Salam [3.100] have shown that with this pattern (3.134) of symmetry breaking, the mass scale M needed to explain the observed α/α_s ratio is $M \simeq 10^{4-6}$ GeV. That is, there is no need for an extreme hierarchy of mass scales in the SU_4^4 model. However, the low energy value of $\sin^2\theta_W$ comes out too high ($\simeq$0.30) in this model.

The low unification mass in SU_4^4 can be traced back to the embedding of SU_3^c in the chiral color group, which is unbroken for $M^2 > Q^2 > M_A^2$. (Other breaking patterns are considered in [3.100,3.103].) Elias and Rajpoot [3.104] have shown that if SU_4^4 is extended to SU_{2n}^4, $n \geq 3$ (which allows n quark doublets and n (2n - 3) lepton doublets to be incorporated into the basic representations $f_{L,R}$), then a low unification mass (10^6 GeV) is allowed even if SU_{2n}^4 breaks directly to $G_s = SU_3^c \times SU_2 \times U_1$ at M. The predicted values of $\sin^2\theta_W$ are in the experimentally favored range 0.21 to 0.25. (For models which break directly to G_s, M and $\sin^2\theta_W$ are determined by the representation content of the fields. See Section 6.2.)

Fractional Versus Integer Charged Quarks and Baryon Number Violation

Models such as SU_4^4 allow two alternate patterns for the final stage of symmetry breaking [3.94,3.99], which lead respectively to fractionally charged quarks (FCQ) or integer charged quarks (ICQ). In the SU_4^4 model there is a Higgs multiplet C which transforms as (4, 1, 4*, 1). (There are also other multiplets related by discrete symmetries.) The VEV of C is of the form

$$\langle C \rangle = \begin{pmatrix} c_1 & & & 0 \\ & c_1 & & \\ & & c_1 & \\ 0 & & & c_4 \end{pmatrix}, \qquad (3.135)$$

where c_4 generates the W_L mass. Depending on the details of the Higgs potential, c_1 will either be zero or non-zero.

If c_1 is exactly zero, the standard model is recovered, with c_4 (an SU_2 doublet) breaking $SU_3^c \times SU_2 \times U_1$ down to $SU_3^c \times U_1^{EM}$. In this case, the quarks are fractionally charged (with the charge independent of color), the gluons are massless, and quarks and gluons are presumably confined. Also, B and L are conserved in this case, implying that quarks and nucleons are stable. (Extended models will generally allow proton decay.) The model contains lepto-quark X bosons of charge ±2/3 and mass $M_X \sim M \simeq 10^{4-6}$ GeV, associated with the $SU_4 \times SU_4$ color group. The X bosons carry baryon and lepton number from one side of a diagram to another, as in Fig. 3.7, but do not mediate proton decay (they have no diquark coupling). They can, however, mediate apparent FCNC processes such as $K_L \to \mu e$. The phenomenology of the X bosons, of the axial gluons, and of the new weak bosons is described in [3.94,3.96,3.99].

If c_1 is non-zero, on the other hand, the physics is entirely different. In this case SU_3^c is spontaneously broken as a local symmetry (global SU_3^c is unbroken), so the eight gluons acquire a common mass proportional to c_1. This presumably implies that quarks and gluons are not confined. Moreover, the electric charge operator, defined as the generator of U_1^{EM} which leaves $\langle C \rangle$ invariant, no longer commutes with SU_3^c. Rather Q is now a symmetric combination of SU_4^S and SU_4^W generators, which leads to integer charged quarks (ICQ). The charges of the particles in $f_{L,R}$ are now

$$q_{L,R} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & -1 \\ -1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \tag{3.136}$$

Finally, the Higgs field corresponding to c_1 carries non-zero B and L (but zero fermion number F = 3B + L), so $c_1 \neq 0$ implies that B and L (but not F) are spontaneously violated, leading to quark and proton decay. The concrete mechanism for B and L violation is that the lepto-quark X bosons, which now have charges ±1 and 0, will mix with the W bosons with mixing mass $\propto \sqrt{c_1 c_4}$. B and L are now violated by the diagrams in Fig. 3.8. These diagrams differ from the B and L violating diagrams of SU_5 in Fig. 3.3 in two respects: (a) the lepto-quark vertex takes a quark into a lepton, not an anti-lepton; (b) B and L are not changed at the lower vertex (i.e., it is not a diquark or dilepton vertex). Hence, fermion number is conserved.

Quarks can decay in the ICQ model by the diagrams [3.94,3.99] in Fig. 3.9. Diagram 3.9a leads to $q \to \nu$ + mesons, while diagram 3.9b, which may or may not be important depending on the $\pi\bar{q}q$ form factor for off shell quarks, can lead to $q \to \nu$ + mesons or e^- + mesons. For $M_X \sim 10^{4-6}$ GeV, as estimated from the unification mass, the quark lifetime (for a 5-10 GeV quark) is estimated to be [3.94,3.99] $\tau_q \sim 10^{-6} - 10^{-9}$ sec, which is short enough for unconfined quarks to have avoided detection. It is interesting that M_X much larger than the unification estimate would lead to long-lived quarks, which should have already been detected, while M_X much smaller is ruled out by the non-observation of $K_L \to \mu e$ (by a diagram similar to 3.7). For further discussion of the production and decays of unconfined integer charged quarks see [3.94,3.99].

The long lifetime of the proton in the ICQ model is attributed to the assumption that quarks and diquarks are more massive than the proton. Hence, the proton can only decay if all three quarks decay simultaneously, as in Fig. 3.10. The decay is therefore third order in the B and L violating interaction and is enormously suppressed. Possible decay modes include $p \to 3\nu\, \pi^+$, $3\nu\, \pi^+\pi^+\pi^-$, $3\nu\, \pi^+\pi^0\pi^0$, and (if 3.9b is important) $p \to \nu\nu e^-\, \pi^+\pi^+$. Note that there are a minimum of four particles in the final state and that $F = 3B + L$

is conserved. Pati and Salam estimate [3.94,3.99] a proton lifetime τ_p of 10^{29}-10^{34} years.

Hence, the proton is either stable or unstable in the SU_4^4 model, depending on whether quarks have fractional or integer charge. (If SU_4^4 is embedded in a larger theory, such as SU_{32}, there may also be fermion number violating decay modes such as $p \rightarrow e^+ \pi^0$. These can occur for either ICQ or FCQ.) Perhaps the most exciting feature of the model is that reasonable values of α/α_s and τ_p can be achieved with a low (10^{4-6} GeV) unification mass, thus avoiding the need for an extreme mass hierarchy or the associated desert. However, $\sin^2\theta_W$ comes out too high ($\simeq$0.30). The other disadvantage is that the form of the group (the direct product of four factors related by discrete symmetries) and the pattern of light fermion representations and their heavy mirror partners are complicated and apparently rather arbitrary.

4. PROTON DECAY

In this chapter I discuss the general question of baryon number conservation, in particular the question of the stability of the proton and bound neutrons. Section 4.1 contains a few general comments on baryon number conservation. Section 4.2 is a brief review of the current experimental limits on the proton lifetime and of future experiments. The rest of the chapter deals with the theoretical predictions for proton decay. The first suggestion that protons could decay because of the new interactions in grand unified theories was by Pati and Salam [3.99], within the context of semi-simple gauge groups. However, in Section 4.3 I will concentrate on the prediction for the proton lifetime and branching ratios in the SU_5 model. There are two distinct theoretical problems to be considered: the mass M_X of the leptoquark bosons responsible for nucleon decay must be estimated from the mass at which the coupling constants α, α_s, and $\sin^2\theta_W$ are unified (Section 4.3.1), and the proton lifetime and branching ratios must be estimated, the former depending strongly on M_X (Section 4.3.2). Section 4.4 discusses baryon number violation from a more general viewpoint, including a consideration of a more general class of grand unified theories and theories in which the proton is stable.

There are several other reviews of various aspects of proton decay. References [4.1] and [4.2] review the experimental limits on τ_p. [4.1] and [4.3] describe the general theoretical ideas in a non-technical sense. References [4.4] - [4.6] are technical reviews of proton decay, mainly in the SU_5 model.

4.1 Baryon Number

The absolute or approximate stability of protons and some bound neutrons is usually attributed to the exact or approximate conservation of baryon number (B). In order to discuss the possibility of baryon number violation, it is useful to contrast it with another conservation law that is seldom questioned, that of electric charge. The difference, of course, is that electric charge is associated with a gauge symmetry. There is no acceptable way to violate electric charge without also giving mass to the photon [4.7]. As the experimental limit on the photon mass is extremely stringent ($m_\gamma < 6\times10^{-22}$ MeV [2.34]), it is reasonable to assume that Q is exactly conserved.

Baryon number does not appear to be associated with a gauge symmetry, however. If it were, the associated long range force would couple to baryon number and not mass. It was pointed out by Lee and Yang [4.8] that such a force would generate an apparent difference between gravitational and inertial mass, which could be detected in the Eötvös experiment [4.9]. From the present limits [4.9] one can conclude that the upper limit on the coupling of such a baryonic photon is $\alpha_B \lesssim 10^{-9}\ G_N\ m_p^2 \simeq 6\times10^{-48}$. It therefore seems very likely that baryon number is not associated with an unbroken gauge symmetry.

Therefore, if baryon number is exactly conserved it must be because of an unbroken global symmetry. (This is the case in the standard $SU_3^c\times SU_2\times U_1$ model.) Unlike gauge symmetries, however, it is easy to write down interactions which violate the quantum number by a small amount without causing any other difficulties. The validity of baryon number conservation must therefore be considered an experimental question [4.10].

The known interactions are not likely to violate baryon number conservation at an experimentally observable rate. The time scale of the weak interactions is of order 10^{-10} sec, and that of the strong and electromagnetic interactions is even less. The present lower limit of $\tau_p > 10^{30}$ years therefore suggests that these interactions conserve B, although one cannot rule out the possibility of very tiny B violating components. ('t Hooft has observed [4.11] that baryon and lepton number will be violated by vacuum tunneling effects in the ordinary weak interactions. For example, the reactions $p + n \to e^+ \bar{\nu}_\mu$ and $p + n \to \mu^+ \bar{\nu}_e$ will occur in the $SU_2 \times U_1$ model. However, this source of baryon number violation is negligible in practice because decay rates and cross sections are proportional to $\exp[-4\pi \sin^2\theta_W/\alpha] \simeq \exp[-400]$!) Gravitational interactions, on the other hand, are of order 10^{-33} of typical weak amplitudes, so that even if gravity somehow violated baryon number it would lead to an unobservably long ($>10^{50}$ yr) proton lifetime.

It is therefore likely that if the proton does decay at an observable rate it is because of a new interaction. This is exactly what occurs in most grand unified theories. In most such theories baryon number is explicitly violated by the gauge and Yukawa couplings of fermions to the new bosons in the theory. The long lifetime of the proton is attributed to the fact that the relevant leptoquark and Higgs bosons are superheavy.

4.2 Present Limits and Planned Experiments

There are two basic techniques for searching for nucleon decay:
(a) One can detect the residual nucleus that remains after a nucleon within the nucleus decays. Experiments of this type, of which there are nuclear,

geochemical, and radiochemical varieties, are insensitive to the particular decay mode of the nucleon. (b) One can attempt to detect the nucleon decay products. Experiments in this class cannot be sensitive to all possible decay modes. However, they can utilize much larger quantities of matter as nucleon sources and they can generally reduce background levels far below those in the first class of experiments. The various experiments of both types and their results have recently been reviewed by Goldhaber, Langacker, and Slansky [4.1] and by Reines and Schultz [4.2], so I will concentrate here on the most recent experiments and on plans for the future.

One experiment by Reines and collaborators [4.12,4.13] utilized a 20 ton array of CH_2 liquid scintillation detectors located 3200 m underground in a gold mine near Johannesburg, South Africa. (Experiments of this type must be performed far underground to reduce the background from cosmic ray produced muons.) The detectors were sensitive to muons that stop and decay in the detector. These muons could be produced by the decays of protons or neutrons in the surrounding rock. The muons could either be produced directly in the decay or could be the secondary decay products of a π^+ (a π^- will almost always be absorbed or charge exchange before it can decay, while a several hundred MeV π^+ has a probability of $\approx 1/3$ of decaying). During 67 ton-years of running (1965-1974), six muons were observed. This rate was consistent with the number of background events associated with cosmic ray produced neutrinos which interact in the detector or the surrounding rock. Therefore, Reines et al. interpreted their results as an upper limit on the decay rates for $p \to \mu^\pm + X$ and $n \to \mu^\pm + X$. They found:

$$\begin{aligned} \tau(p \to \mu^\pm + X) &> 3 \times 10^{30}\ \mathrm{yr} \\ \tau(n \to \mu^\pm + X) &> 3 \times 10^{30}\ \mathrm{yr}\ . \end{aligned} \tag{4.1}$$

A translation of this result into a total lifetime limit is model dependent. Learned, Reines, and Soni [4.13] estimated that for the branching ratios expected in common models (e.g., SU_5 or the Pati-Salam model) approximately 15% of the nucleon decays would result in a muon. This is an average of a large fraction for proton decays and a small fraction for neutron decays. Equation (4.1) then implies $\tau_p > 10^{30}$ yr.

An ongoing experiment by a Pennsylvania-Brookhaven group [4.14] involves a series of water Čerenkov detectors located 1600 m underground in the Homestake gold mine in South Dakota (the site of the solar neutrino experiments). This experiment searches for muons produced inside the detector, especially those which are moving upwards. (The atmospheric muon background is much more severe than for the much deeper South African experiment.) The experimenters have made a detailed estimate that, for the SU_5 branching ratios, approximately 27% of the proton decays should ultimately produce a muon detectable in their apparatus. Most of these muons are from π^+ or kaon decay. The estimates include the probability that a π^+ will emerge from the oxygen or hydrogen nucleus [4.15] for each decay mode and the probability that the π^+ will decay before being absorbed. Assuming this 27% figure, their current lifetime limit is $\tau_p > 2 \times 10^{30}$ yr. By increasing the volume of the detector (it is currently ≈150 tons) and improving the electronics they should ultimately be sensitive to lifetimes of $\approx 10^{31}$ yr.

A number of other experiments are in various stages of running, construction, or planning. These involve groups from Tata, Irvine-Michigan-Brookhaven (IMB), Harvard-Purdue-Wisconsin (HPW), Minnesota, Frascati-Milano-Torino, and Saclay. Most of these experiments will attempt to identify all of the final particles for some appropriate decay modes. They can therefore yield more detailed information than the experiments that search only for muons. On the other hand, they are sensitive to fewer possible modes.

An important figure of merit is that one ton of matter contains about 6×10^{29} nucleons. Thus, if one makes the rather optimistic assumption of 100% detection efficiency, experiments employing 100 or 10,000 tons of matter as sources of nucleons would be sensitive to nucleon lifetimes of $\simeq 6 \times 10^{30}$ yr and 6×10^{32} yr, respectively. (This is for ten events per year.) Two of the projected experiments, IMB and HPW, will utilize very large quantities ($\simeq$6000 and 1000 tons, respectively) of water as their nucleon source. In each case the water will be surrounded or interspersed with phototubes ($\simeq$2400 for the IMB experiment), which will detect Čerenkov light produced by electromagnetic decay products of the nucleon. For example, $p \rightarrow e^+\pi^0$ would produce three cones of Čerenkov light, while $n \rightarrow e^+\pi^-$ would produce two. (The π^- cone will be distorted by rescattering effects.) Most of the other experiments will employ smaller (100-1000 ton) and denser detectors.

There are two principal backgrounds for these experiments. The first is the highly penetrating muons produced by cosmic ray interactions in the atmosphere. In order to reduce the muon flux to manageable levels it is necessary to perform the experiments deep underground. Even at the depth (600 m) of the IMB apparatus, however, the muons constitute a serious background. Approximately 10^8 muons/year will pass through the detector, about 1% of which will stop. Fortunately, most of these can be easily recognized and can even be used to calibrate the detector. Nevertheless, the muon background should be a serious complication for most of the projected experiments.

A second background, which cannot be eliminated by placing the detector underground, is due to the interactions of neutrinos produced in the atmosphere. For example, the reaction $\bar{\nu}_e p \rightarrow e^+ n \pi^0$, with the neutron undetected, could simulate the decay $p \rightarrow e^+\pi^0$. The background from such events would equal the true signal if the proton lifetime were 5×10^{30} yr. This background can be greatly reduced by measuring the momenta and energies of the final particles. Even when the appropriate kinematic cuts are applied, however, the background would equal the true signal for a lifetime of $\simeq 3 \times 10^{33}$ yr [4.16], independent of the detector size (one must, of course, include the effects of fermi motion in such estimates). It would therefore be very difficult to improve the limit on the nucleon lifetime to much beyond 3×10^{33} yr by terrestrial experiments.

Background from natural radioactivity is of little importance because of the low energies involved.

These large detectors may have interesting secondary uses, such as detecting very high energy extraterrestrial neutrinos, neutrino oscillations, or the decay products of exotic heavy particles [4.17].

The possibility of detecting $\Delta B = 2$ interactions will be discussed briefly in Section 4.4. It has been pointed out [4.18] that the experimental limit on the antiproton lifetime is only $\tau_{\bar{p}} > 10^{-8}$ sec, but that this can be increased to $\simeq 10^{10}$ sec in $\bar{p}p$ storage ring experiments. Of course, any difference between τ_p and $\tau_{\bar{p}}$ would require a violation of the CPT theorem.

4.3 Theory (Mainly SU_5)

4.3.1 Determination of the Lepto-Quark Mass [4.4-4.6]

For any theory in which the grand unified group G breaks down to $SU_3^c \times SU_2 \times U_1$ at a mass M_X, the three coupling constants g_3, g_2, and g_1 of the properly normalized subgroups should come together at or near M_X. Hence, one has two independent determinations of M_X, based on the observed ratios α_s/α_e and $\sin^2\theta_W = \alpha_e/\alpha_g$ at low energies. That these two determinations give the same M_X can be considered a consistency check on the theory. Other constraints on M_X are the proton lifetime and the ratio m_b/m_τ.

For the Georgi-Glashow model, the basic results of Georgi, Quinn, and Weinberg [3.4] for α/α_s and $\sin^2\theta_W$ in the region $M_W^2 < Q^2 < M_X^2$ are given in (3.46) in the approximations of working to lowest order, treating all thresholds as step functions, and neglecting Higgs bosons. Including the effects of n_H light complex Higgs doublets, these become

$$\begin{aligned}\sin^2\theta_W &= \frac{3}{8}\left[1 - \frac{\alpha}{4\pi}\left(\frac{110 - n_H}{9}\right)\ell n\,\frac{M_X^2}{Q^2}\right]\\ \frac{\alpha}{\alpha_s} &= \frac{3}{8}\left[1 - \frac{\alpha}{2\pi}\left(11 + \frac{n_H}{6}\right)\ell n\,\frac{M_X^2}{Q^2}\right]\end{aligned} \tag{4.2}$$

where α, α_s, and $\sin^2\theta_W$ are evaluated at Q^2. (4.2) is independent of the number F of fermion families.

An early estimate of M_X by BEGN [3.3] utilized (4.2) with:
(a) $\alpha_s = 12\pi/(25\ell n\, Q^2/\Lambda^2)$, with $\Lambda \simeq 300$ MeV, valid in the region below heavy quark (c, t) thresholds; (b) $n_H = 0$; (c) $\alpha(Q^2) \simeq \alpha(0) \simeq 1/137.04$. They found $M_X \simeq 3.7\times 10^{16}$ GeV, implying $\sin^2\theta_W\ (M_W^2) \simeq 0.20$ and $\alpha_5(M_X^2)$ (the value of

$g_5^2/4\pi$ at the unification mass) $\simeq 0.022$. For $\tau_p \simeq M_X^4/(\alpha_5^2\, m_p^5)$, this implies $\tau_p \gtrsim 10^{38}$ yr, much too long to be observable.

However, small corrections to (4.2) imply small changes in $\ell n\, M_X$ but large changes in M_X and therefore in $\tau_p \sim M_X^4$. It was subsequently realized, especially by Ross [4.19], Goldman and Ross [4.20], and Marciano [4.21], that most of the corrections tend to reduce M_X (and τ_p). A great deal of effort has therefore gone into the improvement of Eq. (4.2). I will first describe the relation between M_X and α_s, then the relation between M_X and $\sin^2\theta_W$, and finally the question of consistency between α_s, $\sin^2\theta_W$, m_b/m_τ, and τ_p. It will turn out that M_X can be determined rather reliably in terms of α_s, or more precisely, in terms of Λ (M_X and Λ are roughly proportional in the regions of interest). On the other hand, M_X varies exponentially with $\sin^2\theta_W$, so that small uncertainties in $\sin^2\theta_W$ lead to enormous uncertainties in M_X [4.21,4.4]. It is therefore more appropriate to predict $\sin^2\theta_W$ in terms of M_X. The reader is also referred to several excellent recent reviews of similar topics [4.4-4.6].

The Estimate of M_X from α/α_s

The first serious study of the corrections to the relation (4.2) between M_X and α/α_s was by Ross [4.19] and Goldman and Ross [4.20]. They argued that including the Q^2 dependence of α reduces M_X by a factor of 6, that the effects of fermion and boson thresholds reduce M_X by 3, that two loop contributions to the renormalization group equations reduce M_X by 4, and that including a single Higgs doublet reduces M_X by 2. Altogether, therefore, M_X is reduced by a factor $\gtrsim 100$ from the BEGN estimate, implying that the estimate of τ_p is reduced by 10^8 to the experimentally accessible region $\tau_p \simeq 10^{30}$ yr.

(Subsequent improvements in the treatment of α and α_s produced compensating changes in M_X.) I will now describe these and other issues in more detail.

Thresholds and the Renormalization Scheme

There are now a number of studies [4.4-4.6,4.21-4.24a] of the effective coupling constants that appear to be quite different in their treatments of thresholds. Most of the apparent differences are due to different prescriptions for defining renormalized coupling constants (see Section 2.5.2). Clearly, any prescription can be used if it is utilized consistently (i.e., all of the coupling constants must be defined the same way). In fact, we will see that the different approaches yield remarkably consistent results. Two basic approaches have been used, the symmetric momentum subtraction scheme (MOM) and the modified minimal subtraction ($\overline{\mathrm{MS}}$) scheme.

In the MOM scheme the coupling constant is defined in terms of a specific renormalized Green's function (e.g., a triple gluon vertex) in which all of the external legs are given equal spacelike momenta. This scheme, versions of which have been used by BEGN [3.3], Goldman and Ross [4.20], and Ellis et al. [4.4], has the advantage that the Appelquist – Carazzone decoupling theorem [4.25] applies [4.26]. This means that heavy particles explicitly decouple from the renormalization group equations for momenta small compared to their mass. However, the treatment of fermion [3.15] and boson [4.19] thresholds is very complicated. The renormalization group equations depend in a complicated way on the particle masses and must be integrated numerically in the threshold regions [3.3,4.19-4.20,4.4]. There is also a problem in that one should use a Q^2 dependent gauge parameter in this scheme [4.27].

Ross [4.19] and Goldman and Ross [4.20] have concluded that the proper treatment of the boson and fermion thresholds in this scheme reduces the naive estimate of M_X by a factor of 3. The translation of the strong coupling constant determined from deep inelastic scattering, which is interpreted as $\alpha_{\overline{MS}}$ (see Section 2.5.2), to α_{MOM} increases M_X by ≃3 [4.20].

The $\overline{MS}$ scheme has the enormous advantage that the renormalization group equations are independent of mass, allowing a much simpler treatment of thresholds. However, the decoupling theorem no longer explicitly applies [4.26] because all divergent loops contribute to the coupling constant renormalizations, independent of the mass of the internal particle. A very elegant treatment of this problem has been given by Weinberg [4.28] and applied to the estimate of M_X by Hall [4.22]. Weinberg advocates that below the threshold associated with scalar, vector, or fermi particles one should consider the effective field theory in which the degrees of freedom associated with the heavy particles have been integrated out (in the functional integral). This means that below (above) the threshold, the heavy particle contributions should be omitted from (included in) the renormalization group equations. The gauge couplings below and above the threshold are equal up to finite discontinuities that have been computed by Weinberg [4.28]. His result is

$$g_a(Q^2) = g(Q^2) + \frac{g(Q^2)^3}{96\pi^2}\Big[\mathrm{Tr}(L^2_{as}\ \Lambda\ \ell n\ \frac{\mu}{Q} + 8\ \mathrm{Tr}(L^2_{af}\ \ell n(\sqrt{2}\ m/Q)) \\ - 21\ \mathrm{Tr}\{L^2_{av}(\ell n(M/Q) - 1/21)\}\Big] + O(g^5) \ , \tag{4.3}$$

where $g_a(Q^2)$ is the gauge coupling of the a^{th} subgroup (G_a) of the effective field theory below threshold, $g(Q^2)$ is the gauge coupling of the theory above threshold; μ, m, and M are the mass matrices of the heavy scalars, fermions, and vectors associated with the threshold; L_{as}, L_{af}, and L_{av} are the representation matrices of the heavy particles for any one of the generators of G_a, and $\Lambda = 1 - P$ is a projection operator which excludes the Goldstone bosons. For SU_5 broken to $SU_3^c \times SU_2 \times U_1$ at $M_X = M_Y$, for example, then, assuming for simplicity that there are no heavy fermions or scalars with masses near M_X, (4.3) implies that the couplings g_3, g_2, and g_1 of the effective group should all meet g_5 at $Q \simeq e^{-1/21} M_X \simeq 0.953\, M_X$, as illustrated in Fig. 4.1. Alternately, one can interpret (4.3) to mean that $g_a(M_X^2)$ differs from $g_5(M_X^2)$ by a finite discontinuity computable from (4.3). Similarly, (4.3) states that the gauge couplings of the theory below and above a fermion threshold should meet at $\sqrt{2}\, m_f$. (4.3) treats the threshold effects exactly up to order g^3. An alternate derivation of the same result has been given by Binétruy and Schücker [4.23]. Marciano [4.21] and Chang et al. [4.24] have advocated similar treatments of thresholds, although the latter authors mainly apply their results to an asymptotically free variant of the SU_5 model (described briefly in Section 6.4).

$\alpha(Q^2)$

Equation (4.2) is valid above the W threshold. It is typically evaluated at $Q^2 = M_W^2$ or $Q^2 = 4M_W^2$. However, the value 1/137.04 used for the electromagnetic fine structure constant in the early determinations of M_X is really only appropriate at $Q^2 = 0$. Furthermore, it is defined in terms of the vertex of a photon coupled to on shell electrons. Goldman and Ross [4.20] and Marciano [4.21] pointed out that α in (4.2) should be evaluated at $Q^2 = M_W^2$ (or $4M_W^2$). It turns out that this is the single largest correction to the early value $M_X = 3.7 \times 10^{16}$ GeV.

The basic equation for $\alpha(Q^2)$, still defined in terms of the vertex of a photon with on-shell electrons, is [4.21]

$$\alpha^{-1}(Q^2) - \alpha^{-1}(0) = -\frac{1}{3\pi}\sum_f q_f^2 \,\ln \frac{Q^2}{M_f^2} \tag{4.4}$$

where the sum extends over all fermions of charge q_f with mass $M_f < Q$. For the quarks, M_f is the constituent mass. (4.4) follows easily from (2.238) in the approximation of treating fermion thresholds as step functions. (Goldman and Ross [4.20] obtain a slightly different formula from a different treatment of the thresholds.) Because (4.4) depends (weakly) on the constituent quark masses, Ellis et al. [4.4] have also employed an alternate method (used earlier by Paschos [4.29]) in which $\alpha^{-1}(Q^2)$ is determined by a dispersion integral over the measured cross section for $e^+e^- \to$ hadrons.

In addition, α^{-1} defined in terms of an on-shell vertex function must be converted into the α^{-1} relevant to the MOM or $\overline{\mathrm{MS}}$ subtraction schemes. The translations are [4.20]

$$\alpha^{-1}_{\mathrm{MOM}}(4M_W^2) = \alpha^{-1}(4M_W^2) + 0.60 \tag{4.5}$$

and [4.23,4.30]

$$\alpha^{-1}_{\overline{\mathrm{MS}}} = \alpha^{-1} - 0.83 \ . \tag{4.6}$$

Finally, Hall [4.22] has calculated $\alpha^{-1}_{\overline{\mathrm{MS}}}$ at $Q = 0.95\ M_W$ using the effective field theory method.

The results of all of these calculations, which are in good agreement with each other, are shown in Table 4.1. The result of using the correct value for α in (4.2) is to reduce the value obtained for M_X by $\simeq 10$. (The early estimate of a factor 6 was based on a preliminary value for $\alpha^{-1}(4M_W^2)$ [4.20].)

Table 4.1 Various values for $\alpha^{-1}(Q^2)$.

Marciano [4.21]	Step Function	$\alpha^{-1}(M_W^2) = 128.5 \pm 0.5$
	$M_u = M_d = M_s = 1.0$ GeV	
	$M_c = M_b/3 = M_t/9 = 1.5$	
Goldman and Ross [4.20]	Modified Step Function	$\alpha^{-1}(4M_W^2) = 128.8$
	$M_u = M_d = .3$	$\alpha_{MOM}^{-1}(4M_W^2) = 129.4$
	$M_s = .5$, $M_c = 1.5$	
	$M_b = 5$, $M_t = 25$	
Ellis et al. [4.4]	Dispersion Relation	$\alpha^{-1}(M_W^2) = 128.64 \pm 0.42$
		$\alpha^{-1}(4M_W^2) = 127.43 \pm 0.47$
		$\alpha_{\overline{MS}}^{-1}(M_W^2) = 127.81 \pm 0.42$
		$\alpha_{MOM}^{-1}(4M_W^2) = 128.03 \pm 0.47$
Hall [4.22]	Effective Field Theory	$\alpha_{\overline{MS}}^{-1}((.95\ M_W)^2) = 128.2$

Two Loop Contributions to the Renormalization Group Equations

Goldman and Ross [4.20] included the two loop contributions to the renormalization group equations in their analysis. Equations (2.243) are replaced by

$$\frac{d\ \alpha_a^{-1}}{d\ \ell n\ Q^2} = \frac{\beta_0^a}{4\pi} + \sum_{b=1}^{3} \frac{\beta_1^{ab}\ \alpha_b}{(4\pi)^2}\ , \tag{4.7}$$

where $\beta_0^1 = -4F/3$, $\beta_0^2 = (22 - 4F)/3$, $\beta_0^3 = (11 - 4F/3)$, and

$$\beta_1^{ab} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{136}{3} & 0 \\ 0 & 0 & 102 \end{pmatrix} - F \begin{pmatrix} \frac{19}{15} & \frac{3}{5} & \frac{44}{15} \\ \frac{1}{5} & \frac{49}{3} & 4 \\ \frac{11}{30} & \frac{3}{2} & \frac{76}{3} \end{pmatrix} , \qquad (4.8)$$

where F is the number of families. (4.7) is typically solved by iteration, with the lowest order expression for α_a substituted on the right-hand side. (See Ref. [4.29], for example, for an explicit expression for the leading correction.) For $F = 3$, the two loop terms lower the prediction for M_X by a factor of 4 [4.20]. Of course, if one utilizes (4.7) then consistency requires that one use a value for α_s that includes two loop corrections.

Effects of Additional Light Particles

Including the effects of a single complex doublet of light ($< M_W$) Higgs particles lower the estimate of M_X by a factor of 1.8 [4.20,3.3]. Each additional complex doublet lowers M_X by approximately the same factor. Hence, each additional light Higgs doublet reduces the estimate of τ_p by an order of magnitude. This will be a serious constraint on n_H.

The extrapolation of α and α_s from low energies to M_W^2 depends on the fermion mass spectrum [4.27]. Fortunately, the sensitivity to the t quark mass is small [4.4,4.22-4.23]: M_X decreases by $\lesssim 2\%$ as m_t is increased from 15 to 50 GeV. This is because the effects of m_t on α and α_s compensate to leading order [4.23]. A more serious uncertainty involves the possibility of more light ($m \lesssim M_W$) families of fermions (although additional families are disfavored because of cosmological constraints (Section 2.4.4) and the m_b/m_τ ratio (Section 3.3.1)). The leading order Eqs. (4.2) are independent of the number F of families, but the two loop corrections introduce an F dependence. The result is [4.20,4.4] that M_X increases by a factor of $\simeq 1.8$ for each additional light generation.

Ellis et al. [4.4] have discussed the possibility of combining technicolor (Sections 2.5.3 and 6.9) with SU_5 in a group $SU_5 \times G_{TC}$. If one assumes that there are F_{TC} technicolor families, each of which transforms as a $5^* + 10$ under SU_5, then the TC families will increase M_X by $\simeq 2 \times 2^{F_{TC}}$, where the first factor is due to the omission of the usual Higgs doublet. The uncertainties associated with the strong TC interactions (e.g., the effects of resonances) are at least a factor $(1.5)^{\pm F_{TC}}$.

Effects of Heavy Particles

Several authors have discussed the effects of superheavy colored scalar fields. Ellis, Gaillard, and Nanopoulos [4.31] have argued that the effects of the color triplet in the 5 will not be important for proton decay unless the Higgs mass is $\lesssim 10^{10}$-10^{11} GeV, but that masses this low compared to M_X are unlikely (e.g., see [3.3] and Eq. (3.56)), especially when radiative corrections to the Higgs potential are included [4.32]. Goldman and Ross [4.20] estimated that a 10^{10} GeV Higgs color triplet would increase M_X by $\simeq 1.2$. Hall [4.22], using the effective field theory method, argues that varying the Higgs mass from 10^{-3} M_X to 10^{+3} M_X leads to an uncertainty of 1.5 in M_X. Cook et al. [4.33-4.34] have pointed out that although the effects of heavy colored scalars in a 5 have little effect on M_X, the introduction of a 45 of Higgs would lead to a large uncertainty of a factor of 3 (in either direction) in M_X.

Ellis et al. [4.4] have also discussed the uncertainties in M_X due to superheavy fermions. For example, each $5+5^*$ family of superheavy fermions (such as is expected for the embedding $SU_5 \subset SO_{10} \subset E_6$) introduces an uncertainty of a factor of three in M_X.

Results

Table 4.2 summarizes the results of the various analyses of M_X as a function of $\Lambda_{\overline{MS}}$. Each of the authors concludes that M_X and $\Lambda_{\overline{MS}}$ are approximately linearly related in the region of interest. Ellis et al. [4.4] have taken $\Lambda_{\overline{MS}} \simeq 0.4$ GeV, with an uncertainty of a factor $(1.5)^{\pm 1}$. (See also Section 2.5.2.) The values of M_X for $\Lambda_{\overline{MS}} = 0.4$ GeV are also shown. The reader is referred to the original papers for the authors' estimates of uncertainties due to higher order corrections, the treatment of thresholds, the t quark mass, and the masses of heavy Higgs bosons. A reasonable consensus is that these effects lead to a further uncertainty of $(1.5)^{\pm 1}$ in M_X.

The results in Table 4.2 are in excellent agreement with each other (within the $(1.5)^{\pm 1}$ uncertainty). In particular, the calculations based on the MOM and $\overline{MS}$ schemes agree to within 50% in M_X and to within $\simeq 1\%$ in $\ln(M_X/M_W)$. This is very encouraging in that the two schemes involve a very different ordering of the perturbation theory [4.22].

For concreteness I will take

$$M_X = 15 \times (1.5)^{\pm 1} \times 10^{14} \; \Lambda_{\overline{MS}} \tag{4.9}$$

with $\Lambda_{\overline{MS}} = 0.4 \times (1.5)^{\pm 1}$ GeV. For $\Lambda_{\overline{MS}} = 0.4$ GeV, (4.9) implies

$$M_X = 6 \times (1.5)^{\pm 1} \times 10^{14} \text{ GeV} \quad . \tag{4.10}$$

Table 4.2 Estimates of M_X for $n_H = 1$, $F = 3$. The estimates all agree to within the 50% uncertainty due to higher order terms, t quark mass, masses of heavy Higgs particles, etc.

Authors and Method	M_X	M_X (GeV) for $\Lambda_{\overline{MS}} = .4$ GeV
MOM		
Goldman and Ross [4.20]	$1.08 \times 10^{15}\ \Lambda_{\overline{MS}}^{0.98}$	4.4×10^{14}
Ellis et al. [4.4]	$1.35 \times 10^{15}\ \Lambda_{\overline{MS}}^{1.01}$	5.4×10^{14}
$\overline{MS}$		
Binétruy and Schücker [4.23]	$1.67 \times 10^{15}\ \Lambda_{\overline{MS}}^{1.03}$	6.5×10^{14}
Hall [4.22]	$1.5 \times 10^{15}\ \Lambda_{\overline{MS}}$	6.0×10^{14}
Marciano [4.21]	$1.6 \times 10^{15}\ \Lambda_{\overline{MS}}$	6.3×10^{14}
Ellis et al. [4.4]	Same as Binétruy and Schücker	

(4.9) is for one light Higgs doublet and $F = 3$. M_X decreases (increases) by a factor of $\simeq 1.8$ for each additional Higgs doublet (fermion family). Discussions of the effects of more exotic fields (e.g., 45's of Higgs, technicolor, or superheavy fermions) are cited in the text.

$\alpha_5(M_X^2)$ is not significantly changed by the corrections to (4.2). Goldman and Ross [4.20] obtain $\alpha_5(M_X^2) = 0.0244 \pm 0.0002$.

The Relation Between $\sin^2\theta_W$, Λ, and M_X

The first serious improvement in the formula (4.2) for $\sin^2\theta_W$ was by Marciano [4.21], who included the Q^2 dependence of α and the effects of a

Higgs doublet. Marciano also argued that the parameter measured in neutral current experiments should coincide with $\sin^2\theta_W(M_W^2)$ to within 0.01. Subsequently, the two-loop corrections to the renormalization group equations were incorporated by Paschos [4.29], Mahanthappa and Sher [4.34-4.35], Marciano [4.6], and others [4.20,4.22-4.23,4.36]. All of the calculations are in good agreement with each other, except for that of Paschos which was evaluated at $Q = 10$ GeV rather than M_W. In Fig. 4.2 I plot the relation obtained by Marciano between $\sin^2\theta_W(M_X)$, M_X, and $\Lambda_{\overline{MS}}$. (For a given $\Lambda_{\overline{MS}}$, the results of Refs. [4.20] and [4.22] are almost identical, while values of $\sin^2\theta_W$ about 1 and 2% lower are found in [4.23] and [4.34].) For fixed M_X, the prediction for $\sin^2\theta_W$ increases by $\simeq 0.0015$ for each additional light Higgs doublet [4.21] and decreases by $\simeq 0.01$ for each extra family of light fermions [4.20].

It is apparent from Fig. 4.2 that $\sin^2\theta_W$ is predicted rather precisely. For $n_H = 1$, $F = 3$ and $M_X = 6 \times (1.5)^{\pm 1} \times 10^{14}$ GeV one has

$$\sin^2\theta_W(M_W^2) = 0.209^{+0.003}_{-0.002} . \tag{4.11}$$

This prediction is much more general than the SU_5 model. It will occur for any theory with no exotic fermions for which the unification group G is broken directly to $SU_3^c \times SU_2 \times U_1$ at M_X (see Section 6.2).

The prediction (4.11) for $\sin^2\theta_W(M_W^2)$ is in reasonable agreement with the experimental value [2.70] $\sin^2\theta_W = 0.229 \pm 0.009(\pm 0.005)$ (Eq. 2.188), although it is slightly low. Several authors [4.21,4.34-4.35] have emphasized that $\sin^2\theta_W(M_W)$ is not quite the same quantity as the $\sin^2\theta_W$ determined in the phenomenological analyses, because the latter do not include the effects of the weak and electromagnetic radiative corrections. (Recall Marciano's early estimate [4.21] that the two quantities could differ by as much as 0.01.) Recent estimates [4.37] of the radiative corrections to the neutral current deep inelastic cross section indicate that they are small. (The large effects cited in [4.4] were based on an early and incorrect calculation.) However, the most accurate measurements of $\sin^2\theta_W$ are from the neutral current to charged current cross section ratios in deep inelastic neutrino scattering and from weak-electromagnetic interference in the SLAC eD asymmetry experiment [4.38]. At the time of this writing, the radiative corrections have not been computed in entirety for either process. For example, radiative corrections to the charged current processes could modify the true value of ρ in (2.200), presumably within the range allowed in (2.201), leading to a different $\sin^2\theta_W$. Preliminary indications are that these effects will be very tiny, however [4.39].

Consistency of α/α_s, $\sin^2\theta_W$, m_b/m_τ, and τ_p

We have seen that for $n_H = 1$ and $F = 3$, $\Lambda_{\overline{MS}} = 0.4$ GeV implies $M_X = 6\times(1.5)^{\pm 1}\times 10^{14}$ GeV and $\sin^2\theta_W(M_W^2) = 0.209^{+0.003}_{-0.002}$. As discussed in Section 3.3.1, Nanopoulos and Ross [3.16] have shown that the asymptotic prediction $m_b = m_\tau$ is renormalized (under similar assumptions) down to $m_b \lesssim 6$ GeV. Also, this range of M_X will imply a proton lifetime in the acceptable and interesting range $\tau_p \simeq 10^{29}$-10^{33} yr. These results are all roughly consistent with each other and with the experimental data, but I would like to comment briefly on the possibility of obtaining a larger value of $\sin^2\theta_W \simeq 0.23$ by modifying some of the assumptions.

The simplest possibility would be to decrease the unification mass M_X. However, we see from Fig. 4.3 that this would require an unacceptably small

M_X: an increase of 0.01 in $\sin^2\theta_W$ requires a decrease of $\simeq 5.7$ in M_X and $\simeq 1050$ in τ_p. $\sin^2\theta_W = 0.23$ would require $M_X \simeq 1.7 \times 10^{13}$ GeV and $\tau_p \simeq 10^{24}$-10^{26} yr, which is clearly ruled out. One possible loophole is that there are ways to modify the standard SU_5 model to obtain a longer proton lifetime for fixed M_X, as will be described in Section 4.4. Even without the τ_p constraint, however, $M_X = 1.7 \times 10^{13}$ GeV implies $\Lambda_{\overline{MS}} \simeq 11$ MeV, which is too small. I note in passing, however, that a value of $M_X(\Lambda_{\overline{MS}})$ somewhat smaller than the standard estimates would, if one ignores the implications for τ_p, have the desirable consequences of lowering somewhat the estimates of m_b and m_s [3.16], as well as increasing $\sin^2\theta_W$.

A second possibility is to increase the number F of light fermion families. For fixed $\Lambda_{\overline{MS}}$, $M_X(\tau_p)$ is increased [4.20,4.4] by $\simeq 1.8(10)$ for each family above $F = 3$, but unfortunately the prediction for $\sin^2\theta_W$ decreases (by $\simeq 0.01$) [4.20]. Also, the renormalized values of m_b and m_s increase to unacceptably large values for $F > 3$ [3.16], although the quark mass predictions can always be modified by including 45 dimensional Higgs representations.

A third possibility is to increase the number n_H of Higgs doublets. Here one must be very careful to specify what is held fixed. For fixed M_X (and τ_p), $\sin^2\theta_W$ increases by $\simeq 0.0015$ for each additional Higgs doublet. Unfortunately, a disquieting number ($n_H \simeq 14$) of Higgs doublets would be required to increase $\sin^2\theta_W$ to 0.23. Moreover, this leads to unacceptably small values for α/α_s. (In fact, the expression (4.2) for α/α_s becomes negative for $n_H > 14$.) Also, m_b increases slightly with increasing n_H [4.40]. The relations between $\sin^2\theta_W$, M_X, n_H, m_b/m_τ are discussed more carefully, in the one loop approximation, by Komatsu [4.40].

Finally, one can attempt to vary n_H and M_X simultaneously. To analyze this, let us return to the one loop equations (4.2). If one requires $\sin^2\theta_W = 0.23$ and $\Lambda = 300$ MeV then (4.2) can be satisfied for $n_H \simeq 7$ and $M_X \simeq 5\times 10^{13}$ GeV for $\alpha^{-1} \simeq 128$. This will again lead to much too short a proton lifetime (especially when higher order effects are included) unless proton decay is somehow suppressed.

In summary then, a value of $\sin^2\theta_W(M_W^2)$ as large as 0.23 is essentially impossible to obtain in the standard SU_5 model. If improved experiments and complete calculations of the necessary radiative corrections indicate that such a large value is needed, then one would have to go to more complicated models in which proton decay is suppressed (Section 4.4), the asymptotic values for $\sin^2\theta_W$ and α/α_s are modified (Section 6.2), there are intermediate thresholds between M_W and M_X (Section 6.2), or the low energy theory is different from $SU_3^c \times SU_2 \times U_1$ (Section 6.2).

4.3.2 Determination of the Proton Lifetime and Branching Ratios

In this section I will discuss the proton lifetime and branching ratios in the SU_5 and, to a limited extent, the SO_{10} models, assuming that colored Higgs particles are sufficiently massive that their effects can be ignored. A somewhat more general discussion is given in Section 4.4.

The Effective Interaction

From (3.39) or (3.116) one can easily write down effective four fermion interactions for baryon number violating processes in the SU_5 and SO_{10} models. If one ignores mixing effects as well as the second and third families, the SU_5 effective interaction is [3.3]

$$\mathscr{L}_{\text{eff}} = \frac{4G}{\sqrt{2}}\left[\left(\varepsilon_{\alpha\beta\gamma}\,\overline{u}_L^{c\gamma}\,\gamma^\mu\,u_L^\beta\right)\left(2\overline{e}_L^{+}\,\gamma_\mu\,d_L^\alpha + \overline{e}_R^{+}\,\gamma_\mu\,d_R^\alpha\right)\right.$$
$$\left. - \left(\varepsilon_{\alpha\beta\gamma}\,\overline{u}_L^{c\gamma}\,\gamma^\mu\,d_L^\beta\right)\left(\overline{\nu}_R^c\,\gamma_\mu\,d_R^\alpha\right)\right] + \text{H.C.}\ , \tag{4.12}$$

where

$$\frac{G}{\sqrt{2}} = \frac{g_5^2}{8M_X^2} \tag{4.13}$$

defines the analogue of the fermi constant of the weak interactions and $M_X = M_Y$ has been assumed. (Some early references have the opposite sign for the e_R^+ term. The sign given above is correct in the interaction basis for the sign conventions in Chapter 3, and also in the mass basis (ignoring mixings) if the e and d masses are generated by a 5 of Higgs.) Equation (4.12) can be derived from (3.39) using the Fierz identity

$$\overline{\psi}_{1L}\,\gamma_\mu\,\psi_{2L}\,\overline{\psi}_{3L}\,\gamma^\mu\,\psi_{4L} = \overline{\psi}_{1L}\,\gamma_\mu\,\psi_{4L}\,\overline{\psi}_{3L}\,\gamma^\mu\,\psi_{2L}\ , \tag{4.14}$$

where the ψ_i are anticommuting fermion fields, and the antisymmetry of the quark fields in the color indices. The identity

$$\overline{\psi}_{1L}\,\gamma_\mu\,\psi_{2L} = -\,\overline{\psi}_{2R}^c\,\gamma_\mu\,\psi_{1R}^c \tag{4.15}$$

is also useful in the following calculations.

From (3.116) one can also derive the effective interaction due to the X' and Y' bosons of the SO_{10} model:

$$\mathscr{L}'_{\text{eff}} = \frac{4G'}{\sqrt{2}}\left[\left(\varepsilon_{\alpha\beta\gamma}\,\overline{u}_R^{c\gamma}\,\gamma^\mu\,d_R^\beta\right)\left(-2\overline{\nu}_L^c\,\gamma_\mu\,d_L^\alpha\right)\right.$$
$$+ \left(\varepsilon_{\alpha\beta\gamma}\,\overline{u}_L^{c\gamma}\,\gamma^\mu\,u_L^\beta\right)\left(\overline{e}_R^{+}\,\gamma_\mu\,d_R^\alpha\right) \tag{4.16}$$
$$\left. - \left(\varepsilon_{\alpha\beta\gamma}\,\overline{u}_L^{c\gamma}\,\gamma^\mu\,d_L^\beta\right)\left(\overline{\nu}_R^c\,\gamma_\mu\,d_R^\alpha\right)\right] + \text{H.C.},\ ,$$

where

$$\frac{G'}{\sqrt{2}} = \frac{g^2}{8M_{X'}^2} \quad , \tag{4.17}$$

with $g = g_5$ and $M_{X'} = M_{Y'}$. The SU_5 singlet ν_L^c may be very massive.

Equations (4.12) and (4.16) are easily generalized to include additional families and mixings between families. For the SU_5 interactions in (3.86), which are valid if all fermion masses are generated by Higgs 5's, the effective interactions relevant to proton decay are [3.22]

$$\begin{aligned} e^{-i\alpha_1} \mathscr{L}_{eff} = \frac{4G}{\sqrt{2}} \Bigg[& \left(\varepsilon_{\alpha\beta\gamma} \bar{u}_L^{c\gamma} \gamma^\mu u_L^\beta \right) \\ & \Bigg\{ \left[\left(1 + c^2 \right) \bar{e}_L^+ + sc\, \bar{\mu}_L^+ \right] \gamma_\mu d_L^\alpha \\ & + \left[\left(1 + s^2 \right) \bar{\mu}_L^+ + sc\, \bar{e}_L^+ \right] \gamma_\mu s_L^\alpha \\ & + \bar{e}_R^+ \gamma_\mu d_R^\alpha + \bar{\mu}_R^+ \gamma_\mu s_R^\alpha \Bigg\} \\ & - \left[\varepsilon_{\alpha\beta\gamma} \bar{u}_L^{c\gamma} \gamma^\mu \left(c\, d_L^\beta + s\, s_L^\beta \right) \right] \\ & \left[\bar{\nu}_{eR}^c \gamma_\mu d_R^\alpha + \bar{\nu}_{\mu R}^c \gamma_\mu s_R^\alpha \right] \Bigg] \\ & + \text{H.C.} \quad , \end{aligned} \tag{4.18}$$

where $c = \cos\theta_c$, $s = \sin\theta_c$, and couplings to the third family have been neglected.

Symmetry Principles

Many important results follow directly from the effective Lagrangians in (4.12), (4.16), or (4.18).

(a) Although B and L are violated, the combination B - L is conserved [4.41]. Thus, the decays $p \to e^+X$ or $p \to \nu^c X$ are allowed, while $p \to e^-X$ or $p \to \nu X$ are forbidden.

(b) $\Delta S = 0$ or $\Delta S = -\Delta B$. Hence, $p \to \nu^c \pi^+$ and $p \to \nu^c K^+$ are allowed, while $n \to e^+ K^-$ and $p \to \nu^c K^- \pi^+ \pi^+$ are forbidden [3.59].

These results are much more general than the SU_5 and SO_{10} models [4.41-4.43]. Weinberg [4.41] classified the $G_s = SU_3^c \times SU_2 \times U_1$ quantum numbers of the bosons which could couple to fermions and mediate nucleon decay. He found that the only possible bosons are vectors with the quantum numbers of (X, Y) or (X', Y') and three types of color triplet scalars. These are SU_2 singlets with electric charge -1/3 (such as H^α) or -4/3, and an SU_2 triplet with charges 2/3, -1/3, and -4/3. These all satisfy $\Delta B = \Delta L$ and $\Delta S/\Delta B = -1, 0$, so these selection rules must be respected at tree level in any theory for which M_W/M_X is so small that mixing between bosons (such as X_s-X' mixing in SO_{10}) can be ignored. Weinberg [4.42] and Wilczek and Zee [4.43] then extended this argument to all orders. They showed that the leading contributions (in $1/M_X$) to $\mathscr{L}_{\text{eff}}$ will be G_s invariant four fermion operators ($SU_2 \times U_1$ violating effects will be suppressed by powers of M_W/M_X). They have shown that there are only six such operators (plus their adjoints) involving the fifteen fields in a single family, all of which satisfy $\Delta B = \Delta L$ and $\Delta S/\Delta B = 0, -1$. (A simplified derivation of the $\Delta B = \Delta L$ rule has been given by Lipkin [4.44].) Furthermore, only two of these operators, O_1 and O_2, can be generated by the exchange of superheavy vector bosons (the others are scalar and tensor). They are

$$\begin{aligned} O_1 &= O_{e_R^+} + O_{\nu_R^c} \\ O_2 &= O_{e_L^+} \quad , \end{aligned} \tag{4.19}$$

where

$$O_{e^+_R} \equiv \left(\varepsilon_{\alpha\beta\gamma}\, \overline{u}^{c\gamma}_L\, \gamma^\mu\, u^\beta_L\right)\left(\overline{e}^+_R\, \gamma_\mu\, d^\alpha_R\right)$$

$$O_{\nu^c_R} \equiv \left(-\varepsilon_{\alpha\beta\gamma}\, \overline{u}^{c\gamma}_L\, \gamma^\mu\, d^\beta_L\right)\left(\overline{\nu}^c_R\, \gamma_\mu\, d^\alpha_R\right) \tag{4.20}$$

$$O_{e^+_L} \equiv \left(\varepsilon_{\alpha\beta\gamma}\, \overline{u}^{c\gamma}_L\, \gamma^\mu\, u^\beta_L\right)\left(\overline{e}^+_L\, \gamma_\mu\, d^\alpha_L\right) .$$

The $SU^c_3 \times U_1$ invariance of O_1 and O_2 is obvious. The SU_2 invariance of $O_{e^+_L}$, for example, can be shown by using the Fierz identity to prove that $O_{e^+_L}$ is antisymmetric under $u_L \leftrightarrow d_L$. If one allows a G_s singlet field ν^c_L, then the G_s invariant operator

$$O_3 \equiv O_{\nu^c_L} \equiv \left(-\varepsilon_{\alpha\beta\gamma}\, \overline{u}^{c\gamma}_R\, \gamma^\mu\, d^\beta_R\right)\left(\overline{\nu}^c_L\, \gamma_\mu\, d^\alpha_L\right) \tag{4.21}$$

is allowed, but this will only be relevant for proton decay if ν^c_L is light.

Any single family theory in which nucleon decay is dominated by the exchange of superheavy vector bosons will therefore have an effective Lagrangian of the form

$$\mathscr{L}_G = \frac{4G_1}{\sqrt{2}}\, O_1 + \frac{4G_2}{\sqrt{2}}\, O_2 + \frac{4G_3}{\sqrt{2}}\, O_3 + \text{H.C.} \tag{4.22}$$

For SU_5, for example,

$$\mathscr{L}_{SU_5} = \frac{4G}{\sqrt{2}}\left[2\, O_{e^+_L} + O_{e^+_R} + O_{\nu^c_R}\right] + \text{H.C.} \quad , \tag{4.23}$$

while for SO_{10}

$$\mathscr{L}_{SO_{10}} = \mathscr{L}_{SU_5} + \frac{4G'}{\sqrt{2}}\left[2\, O_{\nu^c_L} + O_{e^+_R} + O_{\nu^c_R} + \text{H.C.}\right] \tag{4.24}$$

It is convenient to define $r \equiv G_2/G_1$. For SU_5, $r = 2$, up to a small correction associated with $SU_2 \times U_1$ anomalous dimensions [4.43]. For SO_{10},

$$r = \frac{2/M_X^2}{1/M_X^2 + 1/M_{X'}^2} \quad , \tag{4.25}$$

which goes to 2, 0, and 1 in the limits $M_{X'}/M_X \to \infty$, 0, and 1, respectively [4.43]. Many of the symmetry relations listed below depend only on r, which is therefore a useful parameter for distinguishing between models in this class.

In the more general case in which mixings are considered there are many allowed operators. In general, each of the four fermions in O_1, O_2, and O_3 can belong to a different interaction basis family, with a correspondence of family indices for the two terms in O_2. For the special SU_5 case in (4.18), for example, $\mathscr{L}_{eff}$ is still in a form similar to (4.23)(generalized to include a second family), except that mixing angles appear because the fields in (4.18) are mass eigenstates.

Weinberg [4.42] and Wilczek and Zee [4.43] have derived a number of additional conclusions from the form of $\mathscr{L}_G$ (some of which were noted earlier in the special cases of SU_5 and SO_{10} by Machacek [3.59]).

(c) $O_{e_R^+}$, $O_{\nu_R^c}$, $O_{e_L^+}$, and $O_{\nu_L^c}$ all transform as doublets under strong isospin. A similar statement holds for the operators involving μ^+ and ν_μ (but with u and d quarks only). To see this, one can use (4.14) and (4.15) to prove that $O_{e_R^+}$ is antisymmetric under $u_R \leftrightarrow d_R$, for example. From the isodoublet nature of these operators one has [3.59,4.42,4.43] (independent of mixing effects)

$$\begin{aligned} \Gamma(p \to e^+_{L,R}\,\pi^o) &= \tfrac{1}{2}\,\Gamma(n \to e^+_{L,R}\,\pi^-) \\ \Gamma(p \to \nu_R^c\,\pi^+) &= 2\;\Gamma(n \to \nu_R^c\,\pi^o) \quad , \end{aligned} \tag{4.26}$$

with similar relations holding for $\pi \to \rho$ and $e^+ \to \mu^+$. ν^c can be ν_e^c, ν_μ^c, or ν_τ^c.

(d) The hadronic parts of $O_{e_R^+}$ and $O_{\nu_R^c}$ are the two components of an isospin doublet (as are the hadronic components of $O_{e_L^+}$ and $O_{\nu_L^c}$). Hence, if one ignores mixing effects (so that $O_{e_R^+}$ and $O_{\nu_R^c}$ have the same coefficient) one has [3.59,4.42,4.43]

$$\Gamma(n \to e_R^+ \pi^-) = \Gamma(p \to \nu_R^c \pi^+)$$

$$\Gamma(p \to e_R^+ X) = \Gamma(n \to \nu_R^c X) \tag{4.27}$$

$$\Gamma(n \to e_R^+ X) = \Gamma(p \to \nu_R^c X) \quad ,$$

up to corrections of order m_e. π can be replaced by ρ in the first line and X is an inclusive sum on hadronic states.

(e) $O_{e_L^+}$ and $O_{e_R^+}$ are mapped into each other under the parity transformation (the $\bar{u}_L^c \gamma^\mu u_L$ term is a vector under parity). Similarly $O_{\nu_L^c} \leftrightarrow O_{\nu_R^c}$ under P. Hence, the SO_{10} effective Lagrangian would be reflection invariant (up to neutrino mass effects) in the limit $r = 1$ ($M_X = M_{X'}$), which would occur if the unbroken electroweak subgroup of SO_{10} at relatively low energies contains $SU_{2L} \times SU_{2R}$.

The consequence of (e) is that

$$\Gamma(N \to e_L^+ X_n) = r^2 \Gamma(N \to e_R^+ H_n) \quad , \tag{4.28}$$

where N = p or n and H_n is any exclusive or inclusive non-strange final state with definite intrinsic parity. I have assumed r is real in (4.28). Combining (4.26), (4.27), and (4.28), one has [3.59,4.42,4.43]

$$
\begin{aligned}
&\Gamma(p \to e^+ \pi^o) = \tfrac{1}{2}\Gamma(n \to e^+ \pi^-) = \tfrac{1}{2}(1 + r^2)\ \Gamma(p \to \nu_R^c \pi^+) = (1 + r^2)\ \Gamma(n \to \nu_R^c \pi^o) \\
&\Gamma(p \to e^+ X) = (1 + r^2)\ \Gamma(n \to \nu_R^c X) \\
&\Gamma(n \to e^+ X) = (1 + r^2)\ \Gamma(p \to \nu_R^c X) \\
&\Gamma(O^{16} \to e^+ X) = (1 + r^2)\ \Gamma(O^{16} \to \nu^c X) \\
&\Gamma(p \to e^+ X) \geq \tfrac{1}{2}(1 + r^2)\ \Gamma(p \to \nu^c X) \quad ,
\end{aligned}
\tag{4.29}
$$

in the absence of mixing. π can be replaced by ρ in the first relation. $r = 2$ in the SU_5 model, which implies the experimentally desirable result that most nucleon decays are into e^+ rather than ν^c [3.3,4.45,3.59]. In the SO_{10} model, however, $r \leq 2$ so that the neutrino modes are relatively more important [3.59]. The X' and Y' by themselves yield $r = 0$. Additional relations involving semi-inclusive decay rates in which a lepton and meson are detected have been given by Hurlbert and Wilczek [4.46].

From (4.28) it is evident that the positron polarization is [4.42-4.43]

$$
P(N \to e^+ H_n) = \frac{1 - r^2}{1 + r^2} \quad . \tag{4.30}
$$

That is, it is a constant for decays into any non-strange final state (as long as m_e can be neglected). More generally [4.42,4.43],

$$
\begin{aligned}
P(N \to \ell^+ H_n) &= \frac{1 - r_{\ell n}^2}{1 + r_{\ell n}^2} \\
P(N \to \ell^+ H_s) &= \frac{1 - r_{\ell s}^2}{1 + r_{\ell s}^2}
\end{aligned}
\tag{4.31}
$$

up to m_ℓ corrections, where $\ell^+ = e^+$ or μ^+, H_n and H_s represent any non-strange or strange final states, respectively, and $r_{\ell n}$ are the relative coefficients of the $\overline{u^c}\, u\, \overline{\ell}_L\, d_L$ and $\overline{u^c}\, u\, \overline{\ell}_R\, d_R$ operators. A similar definition holds for $r_{\ell s}$, with d replaced by s. For $\mathscr{L}_{eff}$ in (4.18), for example, the polarizations of directly produced muons are

$$P(N \to \mu^+ H_n) = -1$$

$$P(N \to \mu^+ H_s) = \frac{1 - (1+s^2)^2}{1 + (1+s^2)^2} \simeq -0.05 \qquad (4.32)$$

$$\overrightarrow{\sin\theta_c = 0} = 0 \quad .$$

(f) There are other relations that are specific to the SU_5 mixing scheme in (4.18). For example [3.22] the ratios of direct muons to positrons are predicted to be

$$\frac{\Gamma(N \to \mu^+ H_n)}{\Gamma(N \to e^+ H_n)} = \frac{s^2 c^2}{(1+c^2)^2 + 1} = 0.010 \quad , \qquad (4.33)$$

for non-strange hadronic states, and

$$\frac{\Gamma(N \to e^+ H_s)}{\Gamma(N \to \mu^+ H_s)} = \frac{s^2 c^2}{(1+s^2)^2 + 1} = 0.023 \qquad (4.34)$$

for S = 1 final states, up to lepton mass corrections.

Gavela et al. [4.47] have also given a number of relations between the widths into specific hadronic states (neglecting the Cabibbo angle), based on hadronic SU_3 and nonrelativistic SU_6 symmetries.

Anomalous Dimensions

The effective Lagrangians in (4.12), (4.16), or (4.18) are derived in tree approximation from the exchange of a single boson. Higher order corrections should be smaller by powers of α_5. This would indeed be the case for matrix elements in which all momenta are of order M_X. However, for proton decay the relevant momenta are of order $\mu \simeq 1$ GeV. Therefore, higher order diagrams of the type shown in Fig. 4.3 will be enhanced by large factors of $\ln \mu/M_X$ or or $\ln (M_W/M_X)$. BEGN [3.3] showed that the leading contributions of the gluon exchange diagrams could be summed to all orders using renormalization group techniques, with the result that $\mathscr{L}_{eff}$ is multiplied by

$$A_3 = \left[\frac{\alpha_3(\mu^2)}{\alpha_5(M_X^2)}\right]^{\frac{2}{11 - 4F/3}} \tag{4.35}$$

Hence, decay rates are enhanced by $A_3^2 \simeq 5$ (for $F = 3$). More recently, Ellis, Gaillard, and Nanopoulos [3.22] and Wilczek and Zee [4.43] have also calculated the enhancement factors due to $SU_2 \times U_1$ bosons. For $n_H = 1$ they find

$$A_2 = \left[\frac{\alpha_2(M_W^2)}{\alpha_5(M_X^2)}\right]^{\frac{27}{86 - 16F}} \simeq 1.42 \tag{4.36}$$

and

$$A_1 = \begin{cases} \left[\dfrac{\alpha_1(M_W^2)}{\alpha_5(M_X^2)}\right]^{\frac{-33}{6 + 80F}} \simeq 1.05 \text{ for } O_1 \\ \left[\dfrac{\alpha_1(M_W^2)}{\alpha_5(M_X^2)}\right]^{\frac{-69}{6 + 80F}} \simeq 1.11 \text{ for } O_2 \quad , \end{cases} \tag{4.37}$$

where the first line applies to all e_R^+, μ_R^+, and ν_R^c operators and the second to all e_L^+ and μ_L^+ operators. Hence, r (or, more generally, $r_{\ell n}$ and $r_{\ell s}$) is renormalized by a factor of $1.11/1.05 \simeq 1.06$ from its tree level value [4.43]. Rates are enhanced by $A_2^2 A_1^2$; which takes the value $\simeq 2.2$ for O_1 decays and $\simeq 2.5$ for O_2 decays. Abbott and Wise [4.48] have calculated A_2 and A_1 for the other G_S invariant operators relevant to nucleon decay.

Estimates of τ_p and Branching Ratios

More detailed estimates of the branching ratios and of the proton lifetime require model dependent assumptions concerning the initial nucleon and final meson wave functions. The many analyses that have been made [4.49-4.53] fall into two major categories. The first class of models [4.49-4.50], pioneered by BEGN [3.3], combine non-relativistic SU_6 symmetric wave functions with parton model ideas. The second group of calculations [4.51-4.53] employ MIT bag model wave functions [4.53].

In the SU_6-parton models, the initial quarks are treated nonrelativistically. Diagrams analogous to Figs. 3.3-3.4 are evaluated in terms of a phenomenological amplitude $\psi(0)$ for the two interacting quarks to be at the same point. $|\psi(0)|^2$ is estimated from hyperon and Ω decays. The third quark is treated as a spectator. (In all existing analyses the nucleon is treated as three valence quarks with sea contributions neglected.) SU_6 symmetry is assumed for the spin, flavor, and color part of the nucleon wave function. In most cases the total lifetime is estimated by integrating over the final ℓ^c and q^c phase space. (This is a parton model type assumption that the sum over physical hadronic final states can be approximated by a sum over noninteracting $q^c q$ states.) The semi-inclusive branching ratios for (p or n) $\rightarrow \ell^+ X_n$, $\ell^+ X_s$, $\nu_\ell^c X_n$, and $\nu_\ell^c X_s$, where ℓ = e or μ and X_n and X_s are inclusive hadronic final states with strangeness 0 and 1, respectively, follow from $\mathscr{L}_{eff}$ and the SU_6 wave function. Branching ratios into exclusive hadronic states are given, under the assumption that the $q^c q$ pair always form a single meson, by projecting the spin, color, and flavor indices of the $q^c q$ pair onto SU_6 wave functions [3.59]. The various models in this category differ from each other mainly in their treatments of phase space and on the method of projecting the spin of the final anti-quarks.

The various estimates of the lifetimes and branching ratios, for both the SU_6-parton and bag calculations, are listed in Tables 4.3-4.7. It is readily seen that there is a considerable variation in the results obtained. Before discussing the results and uncertainties I will outline the major assumptions and approximations of each paper. I have taken some liberties with the lifetime estimates in Table 4.3. In particular I have renormalized the results of Ref. [4.45,4.20,4.47] so as to use the same $|\psi(0)|^2$ in each case, have added an $(A_2 A_1)^2$ enhancement factor to [4.45] (the existence of which was not realized until later), written all lifetimes in the form $\tau_{p,n} = a_{p,n} M_X^4$ (the M_X^4 behavior is modified only very slightly by the enhancement factors), and omitted some early and obviously inconsistent results.

BEGN [3.3] estimated the lifetime from $\mathscr{L}_{eff}$ in (4.12) ignoring mixing, treating the initial quarks at rest (with constituent mass $m_p/3$), and ignoring the masses of the final fermions. They discussed the value of $|\psi(0)|^2$ (see below).

Jarlskog and Yndurain [4.45] included some additional diagrams omitted in [3.3] and estimated the semi-inclusive branching ratios in more detail. They also estimated the effects of the quark "decay" and three body annihilation diagrams shown in Fig. 4.4 and concluded they are less important than the two body graphs in 3.3-3.4.

Within the approximations in [3.3] and [4.45], the nucleon decay rate is of the form

$$\Gamma \simeq \left(\frac{g_5^2}{M_X^2}\right)^2 |\psi(0)|^2 \, m_{qq}^2 \, |A|^2 \, \lambda \, , \tag{4.38}$$

where $m_{qq}^2 \simeq (2m_p/3)^2$, the energy^2 of the initial qq pair, comes from the spin traces and phase space, $A = A_3 A_2 A_1$ is the anomalous dimension enhancement factor, and λ includes numerical phase space, color, spin, and isospin factors.

Machacek [3.59] calculated the semi-inclusive rates assuming static initial quarks. Phase space was calculated assuming constituent quark masses $M_u = M_d = m_p/3$, $M_s = 500$ MeV. (Small) mixing angles derived from the Georgi-Nanopoulos SO_{10} model [3.53] (which should correspond to (4.18)) were included. Exclusive branching ratios were estimated by projecting the spin of the relativistic ($k/M \sim 3/4$) q^c onto SU_6 meson wave functions. Branching ratios for some limits of the SO_{10} model and some Higgs mediated decays were also given.

Ellis, Gaillard, Nanopoulos, and Rudaz [4.4] have reevaluated the lifetime formula using Machacek's phase space assumptions and also reconsidered the value of $|\psi(0)|^2$.

Goldman and Ross [4.20] argued that m^2_{qq} should be suppressed by 30-40% from $(2m_p/3)^2$ and claimed that the uncertainty in the lifetime due to the treatment of the final quark masses is a factor of 2. The corresponding uncertainties in the semi-inclusive branching ratios are small.

Gavela, Le Yaouanc, Oliver, Pène, and Raynal [4.47] have emphasized the predictions of SU_3 and SU_6 symmetry for the nucleon and meson wave functions. The initial and final quarks and antiquark are treated nonrelativistically in the calculation of amplitudes and for the projection onto SU_6 wave functions. In addition to recovering the appropriate special cases of (4.29) and (4.31) they obtained

$$\begin{aligned} \overline{\Gamma}(p \to e^+ \pi^o) &= 3\, \overline{\Gamma}(p \to e^+ \rho^o) \\ \Gamma(p \to \nu^c_\mu K^+) &= 0 \end{aligned} \qquad (4.39)$$

which follow from SU_6,

$$\overline{\Gamma}(p \to e^+ \pi^o) = 3\,\overline{\Gamma}(p \to e^+ \eta)$$
$$\overline{\Gamma}(p \to \mu^+ K^o) = \frac{4}{5}\,\overline{\Gamma}(p \to e^+ \pi^o) \quad , \tag{4.40}$$

which are SU_3 results, and

$$\Gamma(p \to e^+ \omega) = 9\,\Gamma(p \to e^+ \rho^o) \quad , \tag{4.41}$$

which is a consequence of the quark model with ideal ω - ϕ mixing. SU_5 mixing angles have been neglected in (4.39)-(4.41). The bars over the decay rates indicate that the symmetry relations apply only to the amplitudes. They are broken by phase space effects. Gavela et al. [4.47] estimate the actual branching ratios by putting in the correct phase space for each exclusive channel. The total decay rate is obtained by summing over the exclusive rates (rather than by using the parton type approximation). They estimate a small suppression of $\simeq$0.8 in rate for the pionic modes due to recoil effects (i.e., from the momentum dependence of the pion wave function).

Kane and Karl [4.50] emphasized that the exclusive branching ratios depend sensitively on the momentum of the outgoing antiquark. That is, the projection of the q^c spinor onto the SU_6 meson wave function depends on the kinematical assumptions, and some of the discrepancies between calculations can be accounted for by this fact. Kane and Karl present tables in which the relative amplitudes and branching ratios are given for three kinematic models: (a) the static model (NR) in which the q^c is taken at rest. This coincides with the Gavela et al. [4.47] approximation. (b) A recoil model (REC) in which $k/M \sim 3/4$ (analogous to [3.59]). (c) A relativistic model (R) in which the q^c mass is neglected. This should correspond to the bag model calculations discussed below. Kane and Karl take the initial quarks to be at rest, neglect mixings, put in the correct phase space for each channel, and argue that recoil corrections to the pion modes are small.

Before proceeding to the bag models, let me consider the initial qq wave function $|\psi(0)|^2$. There have been several estimates of $|\psi(0)|^2$. Finjord [4.55] obtained $|\psi(0)|^2 \simeq 1.1 \times 10^{-3}$ GeV^3 from Ω^- decay. Schmid [4.56] obtained $\simeq 4.4 \times 10^{-3}$ GeV^3 from S-wave $\Sigma^+ \to p\ \pi^o$. Le Yaouanc et al. [4.57] found $\simeq 11.5 \times 10^{-3}$ from P wave Λ decay. Early estimates of τ_p [3.3,4.45,3.59] used the value $|\psi(0)|^2 \simeq 8 \times 10^{-3}$ GeV^{-3}. However, I will follow Ellis et al [4.58] in renormalizing all results to the lower value 2×10^{-3} GeV^3 which agrees with Ω^- and S wave hyperon decay and is compatible with the bag model (part of the original discrepancy between the bag and SU_6-parton estimates was just due to the value of $|\psi(0)|^2$).

Donoghue [4.51] utilized an MIT bag model. He neglected mixing, put in the correct phase space for each channel, and obtained a total two body decay rate by summing the partial rates. He argued that the large recoil momentum would lead to a suppression of the pionic modes by a factor of three in amplitude (in disagreement with Gavela et al. [4.47] and Kane and Karl [4.50] who subsequently found much smaller effects).

Golowich [4.53], in another bag model, also put in a suppression factor (of 2.5) for the pion amplitudes. He estimated SU_5 branching ratios as well as those for r = 0 and 1. For the SU_5 case the mixing angles in (4.18) were used.

Din, Girardi and Sorba [4.52] employ another bag model. They ignore mixings and study the sensitivity of their results to the assumed quark masses. The results in Table 4.3 are for $M_u = M_d = 0$, $M_s = 280$ MeV. Larger masses can increase τ_p by $\simeq 2$. They assume a suppression of $\simeq 3.5$ in amplitude from Lorentz contraction associated with the recoil of the pion bag.

The results for the p and n lifetime are given in Table 4.3. a_p and a_n are the coefficients in

$$\tau_{p,n}(yr) = a_{p,n}\, M_X\, (Gev)^4 \quad . \tag{4.42}$$

The first three calculations presumably include all hadronic final states, while the last four include only the single meson states. Hence, these calculations include a factor of ρ_p or ρ_n, which are the fractions of two-body final states for p and n decay. For $\rho_{p,n} \simeq 1$ there is a factor of ten discrepancy between the first two bag calculations and the SU_6-parton calculations, despite the similar $|\psi(0)|^2$. The origin of this discrepancy is unknown, though a small part of it is due to the large pion suppression assumed by Donoghue [4.51] and Golowich [4.53]. To add to the confusion, the third bag calculation of Din et al. [4.52] give a result similar to the SU_6-parton calculations.

The major uncertainties in τ_p, given the SU_5 model and the value of M_X, are (a) a factor of $\simeq 2$ uncertainty in $|\psi(0)|^2$; (b) a factor $\simeq 2$ uncertainty from the treatment of quark masses and phase space [4.20,4.52]; (c) a factor of $\lesssim 2$ from possible recoil suppression of the pionic modes; (d) a small uncertainty from the number of families, which affects the anomalous dimensions and the estimate of M_X. For fixed M_X, τ_p decreases by 20-30% for each additional family [4.45,4.52]. Other uncertainties, such as the validity of the parton assumptions, the SU_6 wave functions, the bag model, the fractions of two body decays, and the possibility of calculational errors, are best estimated from the order of magnitude spread of the estimates in Table 4.3. Finally, τ could be increased by large mixing effects if one abandons the minimal Higgs structure (5's and 24's).

From Table 4.3, I conclude (for the SU_5 model with F = 3 and the minimal Higgs structure)

Table 4.3 Lifetime estimates. The first two columns give the coefficients in $\tau_{p,n} = a_{p,n} M_X^4$, where τ is in yr and M_X is in GeV. The last two give the lifetimes for $M_X = 6 \times 10^{14}$ GeV. The first four rows are SU_6-parton calculations with the common value $|\psi(0)|^2 = 1.1 \times 10^{-3}$ GeV^3, which is consistent with the MIT bag wave functions used in the last three rows. ρ_p and ρ_n are the fractions of two body decays for p and n, respectively.

			For $M_X = 6 \times 10^{14}$ GeV	
	$10^{29}\ a_p$	$10^{29}\ a_n$	τ_p (10^{30} yr)	τ_n (10^{30} yr)
JY [4.45]	3.7	4.3	4.8	5.6
EGNR [4.4]	4.8	--	6.2	--
GR [4.20]	2.4	3.7	3.1	4.8
GLOPR [4.47]	5.8 ρ_p	--	7.5 ρ_p	--
D [4.51]	38 ρ_p	50 ρ_n	50 ρ_p	65 ρ_n
G [4.53]	41 ρ_p	48 ρ_n	52 ρ_p	62 ρ_n
DGS [4.52]	2 ρ_p	--	2.6 ρ_p	--

$$\begin{aligned} \tau_p\ (\mathrm{yr}) &= (2.4 - 38) \times 10^{-29}\ M_X^4 \\ \tau_n/\tau_p &\sim 1.1 - 1.5 \quad , \end{aligned} \tag{4.43}$$

where M_X is in GeV (τ_n/τ_p is further discussed below).

Combining this with

$$\begin{aligned} M_X &= 15 \times 10^{14}\ \Lambda_{\overline{MS}} \times (1.5)^{\pm 1} \\ \Lambda_{\overline{MS}}\ (\mathrm{GeV}) &= 0.4 \times (1.5)^{\pm 1} \end{aligned} \tag{4.44}$$

one has

$$\begin{aligned} \tau_p^{(\mathrm{yr})} &= (1.2 - 19) \times 10^{32 \pm 0.7}\ \Lambda_{\overline{MS}}^4 \\ &= (3.1 - 49) \times 10^{30 \pm 1.4} \quad . \end{aligned} \tag{4.45}$$

(4.45) can be rewritten

$$\begin{aligned} \tau_p^{(\mathrm{yr})} &= 4.8 \times 10^{32 \pm 1.3}\ \Lambda_{\overline{MS}}^4 \\ &= 1.2 \times 10^{(31 \pm 2)} \quad . \end{aligned} \tag{4.46}$$

(4.46) differs slightly from the result of Ellis et al. [4.4] because I have renormalized the $|\psi(0)|^2$ used by Gavela et al. [4.47]. The relation (4.46) between τ_p and $\Lambda_{\overline{MS}}$ is shown in Fig. 4.5. (4.46) can be combined with the relation between $\sin^2\theta_W$ and $\Lambda_{\overline{MS}}$ in Fig. 4.2 to obtain τ_p as a function of $\sin^2\theta_W$, as shown in Fig. 4.6. It is seen that the prediction is barely consistent with the experimental value of $\sin^2\theta_W$. However, it should be repeated that radiative corrections have not been included in the determination of $\sin^2\theta_W$.

The predictions for the semi-inclusive branching ratios are shown in Table 4.4. The results are reasonably consistent, with the differences due to the different treatments of phase space (see the discussion of muonic decays

Table 4.4 Estimates of the semi-inclusive branching ratios in the minimal SU_5 model. X_n and X_s are inclusive hadronic states with strangeness 0 and 1, respectively. The two Goldman and Ross (GR) columns use nonrelativistic (NR) and relativistic (R) kinematics. GR combine the $\nu_e^c X_n$ and $\nu_\mu^c X_s$ rates (in the $\nu_e^c X_n$ row).

	Proton Decays				Neutron Decays			
	JY [4.45]	M [3.59]	GR [4.20] NR	GR [4.20] R	JY	M	GR NR	GR R
e^+X_n	80	83	81	80	80	76	72	79
e^+X_s	0	0	--	--	0	0	--	--
μ^+X_n	0	1	--	--	0	1	--	--
μ^+X_s	6	1	11	9	0	0	0	0
$\nu_e^c X_n$	13	13	8	11	19	20	28	21
$\nu_e^c X_s$	0	0	--	--	0	0	--	--
$\nu_\mu^c X_n$	0	<1	--	--	0	1	--	--
$\nu_\mu^c X_s$	1	1	--	--	1	1	--	--

below) and mixing angles. The zero value for $n \to \mu^+ X_s$ is due to the valence quark approximation. The elementary process $uu \to \mu^+ s^c$ could lead to the decay if nucleon sea effects were considered.

The estimates of exclusive branching ratios are shown in Tables 4.5 and 4.6. The NR model of Kane and Karl is in excellent agreement with that of Gavela et al. and their relativistic (R) model is in reasonable agreement with the bag models of Donoghue and Golowich, except for the pionic modes which the latter authors have suppressed by hand. I therefore conclude that the results of Kane and Karl should probably be considered the best available estimates of the two-body branching ratios in the SU_5 model, modulo the possibility that the pionic modes may be somewhat suppressed by recoil effects. The variation of values over their three columns are a reasonable estimate of the uncertainties. Some important general features are that the e^+ modes dominate over the ν^c modes and that the π (except possibly for recoil effects) and ω final states dominate over the η and ρ.

The neutron and proton lifetimes are comparable in the SU_5 model. Various predictions for τ_n/τ_p are given in Table 4.7. The first two entries were obtained by the authors directly from the SU_6-parton calculations. The others are obtained from the branching ratios and (4.29):

$$\begin{aligned} \frac{\tau_n}{\tau_p} &= \frac{B(n \to e^+ M^-)}{B(p \to e^+ M^0)} \frac{\Gamma(p \to e^+ M^0)}{\Gamma(n \to e^+ M^-)} \\ &= \frac{1}{2} \frac{B(n \to e^+ M^-)}{B(p \to e^+ M^0)} \quad , \end{aligned} \tag{4.47}$$

where $M = \pi$ or ρ. The results are all in the range

$$0.8 < \tau_n/\tau_p < 1.5 \quad . \tag{4.48}$$

Table 4.5 Predictions for the branching ratios for proton decay in the SU_5 model. All entries should actually be multiplied by ρ_p, the fraction of two body decays. Columns sometimes do not add up to unity because of roundoff and the omission of minor modes (including 4% estimated by Din et al. for $\pi^o\pi^o e^+$). The static (NR), recoil (REC), and relativistic (R) models of Kane and Karl are described in the text.

Mode	M [3.59]	GYOPR [4.47]	D [4.51]	G [4.53]	DGS [4.52]	KK [4.50] NR	KK [4.50] REC	KK [4.50] R
$e^+\pi^o$	33	37	9	13	31	36	40	38
$e^+\rho^o$	17	2	21	20	21	2	7	11
$e^+\eta$	12	7	3	.1	5	7	1.5	0
$e^+\omega$	22	18	56	46	19	21	25	26
$\nu_e^c\pi^+$	9	15	3	5	11	14	16	15
$\nu_e^c\rho^+$	4	1	8	7	8	1.0	2.6	4
μ^+K^o	.3 - .5	19	--	7	.5	18	8	5
$\nu_\mu^c K^+$	--	0	--	.5	--	0	.2	.6

Table 4.6 Same as Table 4.5, only for neutron decays.

Mode	M [3.59]	GYOPR [4.47]	D [4.51]	G [4.53]	KK [4.50]		
					NR	REC	R
$\nu_e^c\pi^o$	8	8	2	3	8	7	7
$\nu_e^c\rho^o$	4	.5	5	4	.6	1.2	1.8
$\nu_e^c\eta$	3	1.5	1	--	1.5	--	--
$\nu_e^c\omega$	5	3.5	14	10	5	5	5
$e^+\pi^-$	50	74	23	32	79	72	68
$e^+\rho^-$	26	4	55	48	6	12	19
$\nu_\mu^c K^o$	--	10	--	2	1.1	3	0.6

Table 4.7 The ratio τ_n/τ_p of bound neutron to proton lifetime in the SU_5 model.

	τ_n/τ_p
JY [4.45]	1.2
GR [4.20]	1.5
M [3.59]	0.8
GYOPR [4.47]	1.0
D [4.51]	1.3
G [4.53]	1.2
KK (NR) [4.50]	1.1
KK (REC)	0.9
KK (R)	0.9

The relative branching ratios for the non-strange final states can easily be generalized to other models using (4.29). That is, the ratios of $e^+\pi^o/e^+\rho^o/e^+\eta/e^+\omega$ and $\nu_e^c\pi^+/\nu_e^c\rho^+$ for proton decay, and the corresponding ratios for neutron decay, are independent of r, while

$$\frac{\Gamma(p \to e^+M^o)}{\Gamma(p \to \nu_e^c M^+)} = \frac{1 + r^2}{2}$$

$$\frac{\Gamma(n \to e^+M^-)}{\Gamma(n \to \nu_e^c M^o)} = 2(1 + r^2) \tag{4.48}$$

where $M = \pi$ or ρ. These relations are approximately satisfied by all but the first model in Tables 4.5 and 4.6.

Several authors [4.52,4.46,4.59-4.60] have applied soft pion techniques to nucleon decay. Tomozawa [4.59] has estimated $\tau(p \to e^+\pi^o) \simeq 5.6 \times 10^{29}$ yr, which would imply (for a 40% $e^+\pi^o$ branching ratio) $\tau_p = 2 \times 10^{29}$ yr. This is an order of magnitude smaller than any of the estimates in Table 4.3, but it is not clear that soft pion techniques are valid for the very energetic $e^+\pi^o$ decay. The other authors have used soft pion techniques to relate three body amplitudes to two body amplitudes. Din et al. [4.52] conclude that $\Gamma(p \to e^+\pi^o\pi^o)/\Gamma(p \to e^+\pi^o) \simeq 1/7$. Wise, Blankenbecler, and Abbott [4.60], who include Born graphs, PCAC constraints, and final state interactions, find $\Gamma(p \to e^+\pi\pi(I))/\Gamma(p \to e^+\pi^o) \simeq 1/5$ for $I = 0$ and 1.5 for $I = 1$. The $I = 1$ branching ratio includes the ρ, and should be compared with the estimates in Table 4.5.

An important complication is that pions produced by nucleon decay inside a nucleus will have a significant probability of being absorbed, elastically scattered, or charge exchange scattered before they get out of the nucleus. This will significantly affect the probability to produce pions directly or through ρ decay (most η's and ω's will escape the nucleus before decaying).

Sparrow [4.15] has made detailed estimates of these effects. He concluded that in water, for example, for every $p \to e^+\pi^o$ decay (averaged over the two free and eight bound protons), only $\simeq 0.7$ $e^+\pi^o$ pairs will emerge and only $\simeq 0.5$ correlated $e^+\pi^o$ pairs (i.e., with the π^o momentum unaltered) will be produced. Hence, the effective rate for $e^+\pi^o$ is reduced by $\simeq 2$ when kinematic cuts are applied to reduce background. Also, 40% of the correlated pairs that do emerge are from the two free protons. Similarly, the secondary μ^+ rate (from π and ρ decay) is reduced by $\simeq 2.5$ by nuclear effects.

Muons, Strange Particles, and Mixing

There are several sources of muons in nucleon decay: (a) Probably the largest source are the secondary muons from π^+ and kaon decay, as were discussed briefly in Section 4.1 (see also the comment above on nuclear effects).

(b) μ^+ can be directly produced in $\Delta S = 1$ transitions from the quark process $uu \to X \to \mu^+ s^c$, which occurs in the absence of mixing effects. The relevant terms are given for SU_5 in (4.18). There are no additional terms of this type associated with the X' or Y' bosons of the SO_{10} model, as can be seen in (3.116). From Tables 4.4 and 4.5 the SU_5 branching ratio for $p \to \mu^+K^o$ is $\simeq$(5-20)%. (The very small rate found by Machacek [3.59] is apparently the result of taking M_s = 500 MeV, $M_u = M_d = m_p/3$ for all $p \to \mu^+X_s$ modes, which greatly overestimates the phase space suppression for μ^+K^o.) In the valence approximation the $uu \to \mu^+s^c$ transition does not contribute to neutron decay. Decays like $n \to \mu^+K^o\pi^-$ should be allowed at some level, however, by sea effects.

(c) Direct μ^+ can be produced in $\Delta S = 0$ decays (and direct e^+ in $\Delta S = 1$ decays) by mixing effects. In the minimal SU_5 model, with all fermion masses generated by Higgs 5's, the mixing effects are all small and calculable, as in (3.86) and (4.18). One has from (4.33) and (4.34) that the branching ratios for $N \to \mu^+ X_n$ and $N \to e^+ X_s$ are extremely small (mixing was neglected entirely in most of the calculations in Tables 4.4-4.6). The direct μ^+ rate could be enhanced somewhat in the minimal mixing version [4.61] of the SO_{10} model. From (3.116) one sees (with u replaced by A_C u) that

$$\frac{\Gamma(N \to \mu^+ X_n)}{\Gamma(N \to e^+ X_n)} \simeq \sin^2\theta_c \tag{4.49}$$

for Y' exchange (r = 0).

The above statements apply to models with the minimal Higgs schemes (5's for SU_5 or 10's for SO_{10}) so that $M^e = M^\ell$ and $M^\mu = M^{uT}$. As discussed in Chapter 3, more general schemes allow completely arbitrary mixings. One could, for example, have the muon in the same multiplet as the u and d. One therefore loses all predictive power for such ratios as μ/e, $(\Delta S = 1)/(\Delta S = 0)$, and ν^c/e^+. (Implications for τ_p will be considered below.) Unfortunately, the fermion mass spectrum suggests that it may be necessary to go beyond the minimal Higgs structure. One particular modification, first proposed by Georgi and Jarlskog [4.62], leads to mixing effects more complicated than (4.18) but still reasonably small (see Section 6.1.4). DeRújula, Georgi, and Glashow [4.61] have recently discussed the e^+K/μ^+K, $\mu^+\pi/e^+\pi$, and $\ell^+K/\ell^+\pi$ ratios for both the minimal and Georgi-Jarlskog mixing schemes as a function of $M_{X'}/M_X$ in the SO_{10} model.

More generally, Wilczek and Zee [4.43] have introduced the kinship hypothesis, which speculates that any mixing between families will be small (of order θ_c). This means that, up to mixings of $O(\theta_c)$, the light fermions (u, d, u^c, d^c, $e^\pm$, ν_e) are all grouped together in a representation, as are the other families. The kinship hypothesis is satisfied by the minimal and Georgi-Jarlskog schemes. Yoshimura [4.63] has discussed some general conditions under which at least the third family will not mix significantly with the other two.

Ways to Increase τ_p

There are several ways to modify the SU_5 model to allow a longer (or shorter) proton lifetime. As discussed in Section 4.3.1, M_X (and therefore τ_p) can be increased by including additional fermion families or technicolored families. Also, the existence of superheavy fermions or superheavy scalars from a 45 Higgs representation could modify M_X greatly. For fixed M_X the only major way to modify τ_p in the SU_5 model is to allow a violation of the kinship hypothesis (by allowing a 45) [3.21]. If, for example, there were a mixing by angle β between u^c and t^c in (3.35), then τ_p would be increased by $(\cos\beta)^{-2}$. In the extreme (and highly unlikely) case $\beta = \pi/2$, proton decay would be forbidden except for negligible contributions involving virtual heavy particles. Finally, it is possible to make the proton absolutely stable, but this requires a more drastic modification of the theory, as will be discussed in the next section.

Higgs Mediated Decays

Nucleon decay can also be mediated by the exchange of Higgs particles, such as the color triplet H^α in the SU_5 model. Because of the weakness of the Yukawa couplings to light fermions the Higgs mediated decays should be unimportant unless the Higgs mass is $\mu_{H\alpha} < 10^{10-11}$ GeV [4.31], whereas most estimates give $\mu_{H\alpha}$ much larger than this [3.3,4.32]. Branching ratios for

Higgs dominated decays depend on the details of the Yukawa matrix, although one would expect a tendency for decays into the second and third families where kinematically allowed. Machacek [3.59] has estimated the branching ratios for some particular SU_5 and SO_{10} models and has found that ν_τ^c and $S = 1$ modes dominate. To leading order in $1/\mu_H^2$, Higgs mediated decays satisfy [4.42, 4.43] $\Delta B = \Delta L$, $\Delta B/\Delta S = 0, -1$, and $\Delta I = 1/2$ for the $\Delta S = 0$ decays.

4.4 Theory (General)

In this section I will discuss baryon number violation from a somewhat more general viewpoint, including such topics as grand unified theories with a stable proton, B - L conservation, theories with $\Delta B = 2$ interactions, and more general $\Delta B \neq 0$ operators. The baryon asymmetry of the universe is discussed in Chapter 5. Time reversal violation in nucleon decay is mentioned in Section 6.6. Some speculations on the possible breakdown of Lorentz and Poincaré invariance in nucleon decay (i.e., at distance scales of $M_X^{-1} \simeq 10^{-28}$ cm) are given in [4.64].

Models With a Stable Proton

It is possible to construct models in which the proton is absolutely stable [4.65-4.68], although generally at the expense of introducing exotic heavy fermions. Most such models were originally motivated by a desire to avoid the gauge hierarchy problem: if the proton were stable then M_X could be taken to be very small (e.g., 10^3 M_W). It is now clear, however, that a very large M_X is required in most models to unify the coupling constants, so the original motivation is no longer relevant. Nevertheless, the possibility of stabilizing the proton in grand unified theories remains as an interesting option.

We have already encountered some examples of theories with a stable proton in Section 3.4.5. Many of the models based on semi-simple groups have fermion number $F = 3B + L$ as a global symmetry. $B - L$ is a linear combination of a gauge generator and a global generator involving Higgs fields only. It is often the case [3.94,3.95] that, for the SSB pattern leading to fractionally charged quarks, B and L remain as unbroken global symmetry generators.

For models such as SU_5, SO_{10}, E_6, etc., in which quarks, antiquarks, leptons, and antileptons are combined together in irreducible representations, F, B, and L are usually explicitly broken in the Lagrangian.

One way to suppress the proton decay rate would be to modify the mixing matrices, as discussed in the last section, so that the largest amplitudes would be for processes such as $p \rightarrow \tau^+\pi^0$ or $p \rightarrow e^+ b^c d$ that are forbidden by energy conservation. However, it seems very unlikely that the proton decay rate could be suppressed by much more than $\sin^2\theta_c$ by this method.

A number of authors [4.66-4.68] have therefore proposed models in which there is an additional global symmetry $U_{B'}$, for which the generator B' corresponds to baryon number for the light fermions. This symmetry, if not spontaneously broken, forbids proton decay. There is an immediate problem, however. Any additional global symmetry imposed on a gauge theory must commute with the gauge symmetry. It therefore cannot distinguish between quarks and leptons, which are in the same representations. Therefore, B' cannot be just a global symmetry generator of the initial Lagrangian. But B' cannot be a local generator, either, because then it would be associated with a long range force. Therefore, B' must be of the form

$$B' = B_G + B_H \quad , \tag{4.50}$$

where B_G is a gauge generator that distinguishes between quarks and leptons and B_H generates a global symmetry that is imposed on the Lagrangian. It is possible to have a SSB pattern in which B_G and B_H are both spontaneously broken but the linear combination $B' = B_G + B_H$ is preserved as an unbroken global symmetry generator. The single Goldstone boson associated with the orthogonal combination is eaten by the gauge field associated with B_G, so there are no physical Goldstone bosons and no long range force. This mechanism, in which a global symmetry emerges from the simultaneous breaking of a gauge and global symmetry, is due to 't Hooft [4.69]. It was applied to baryon number in a (non-grand unified) toy model by Pais [4.70].

To see a typical consequence of such a symmetry, consider a single family SU_5 model. The usual fermion assignment is

$$\underset{5}{\begin{pmatrix} e^+ & \\ & d \\ \nu_e^c & \end{pmatrix}_R} \qquad \underset{10}{\begin{pmatrix} & u & \\ e^+ & & u^c \\ & d & \end{pmatrix}_L} \qquad (4.51)$$

One can, however, extend SU_5 to $SU_5 \times H$, where the global symmetry H (generated by T_a^a) forbids the Yukawa couplings in (3.64) needed to generate m_u [3.47,3.49]. In fact, the color antitriplet in the 10 cannot even be identified with u^c in this case. Rather, it is a heavy antitriplet field F_α^c, which is forbidden by $B' \equiv \frac{1}{3}\chi \equiv \frac{1}{3}T_\alpha^\alpha$ from mixing with the u^c. The u_L^c and F_R^c must therefore be assigned to other representations, such as [3.47]

$$
\begin{array}{cccc}
 & \begin{pmatrix} e^{+} & & \\ & & d \\ \nu_e^c & & \end{pmatrix}_R & \begin{pmatrix} & u & \\ E^{+} & & F^c \\ & d & \end{pmatrix}_L & \\
N_R^c & & & \\
 & \begin{pmatrix} e^{+} & & \\ & & b \\ N^c & & \end{pmatrix}_L & \begin{pmatrix} & u & \\ E^{+} & & F^c \\ & b & \end{pmatrix}_R & , \\
1 & 5 & 10 &
\end{array}
\qquad (4.52)
$$

where N^c, E^+, and F^c are heavy. This vectorlike model is no longer viable, because of the neutral current structure, but it illustrates the basic idea of stabilizing the proton by replacing the antiquarks and/or leptons in a multiplet with exotic heavy fermions [4.65-4.68]. Fritzsch and Minkowski [4.66] and later authors [4.67-4.68] introduced global symmetries to prevent the heavy particles from mixing with the particles they replace. Yoshimura [3.49] and Langacker, Segrè, and Weldon [3.47] considered vectorlike SU_5 and SU_6 models such as (4.52) in which 3B' is the chromality generator $\chi = T^\alpha_\alpha$, which counts the number of upper minus the number of lower color indices (F^c and u^c in (4.52) have $\chi = +2$ and -1, respectively). Chromality has the advantage that most SSB patterns that conserve SU_3^c also conserve χ. However, it can only be implemented in SU_n models [3.47], and except for one interesting SU_5 model [3.47,4.71] the theory must be vectorlike to avoid a ridiculously large number of fermions. Segrè and Weldon [4.71] have recently reconsidered this one exception, in which each family is placed in a $5_R + 10_L + 10^*_R + 5^*_L + 1_R$. They showed that χ and B - L (and therefore fermion number, which is a linear combination) are conserved, leading to proton stability. However, B and L are violated so that a baryon asymmetry of the universe is still possible (Chapter 5). They have displayed a model in which baryon generation is associated with the decay of exotic fermions at relatively low ($\lesssim 100$ GeV) temperatures. For every baryon ($X = 3$) in the universe there is a long lived heavy ($\simeq 1$ GeV) neutral lepton with $X = -3$.

Other global symmetries have been applied to $SU_{8L} \times SU_{8R}$ [3.97] and E_7 models [4.68]. Gell-Mann, Ramond, and Slansky [3.25] have given a general discussion of the local and global generators B_G and B_H. They have emphasized that B' only coincides with B for the light fermions. There will generally be heavy fermions with "weird" values of B' (such as the $B' = X/3 = -1$ neutral lepton mentioned above).

More General Operators and B - L

In a renormalizable field theory, baryon number violating interactions involving light particles at ordinary momentum scales can be described by an effective Lagrangian of the form [4.25]

$$\mathscr{L}_{eff} = C\, M^{4-d}\, O \quad , \tag{4.53}$$

where O is an operator formed from the light fields, C involves the coupling constants, d is the dimension of O, and M is the mass of the superheavy scalar, vector, or fermi particle which mediates the interaction. (4.53) is true even if the leading contributions to $\mathscr{L}_{eff}$ are loop diagrams.

Wilczek and Zee [4.43,4.72] and Weinberg [4.41,4.42,4.73] have emphasized that one need only consider $G_s = SU_3^c \times SU_2 \times U_1$ invariant operators in (4.53). G_S violation enters through operators involving the Higgs doublet ϕ, which can then be replaced by its VEV. Hence, the effects of G_S violation are suppressed by powers of $\langle\phi\rangle/M \simeq M_W/M$. The above authors and Weldon and Zee [4.74] (who also considered L_e, L_μ, and L_τ violating interactions) have classified the

G_S invariant operators of low dimension (see also the discussions by Pati [3.94] and Marshak and Mohapatra [4.75]).

Baryon number violating operators must involve at least three quark fields to be a color singlet, and at least four fermion fields to be a Lorentz scalar. Hence, the lowest dimensional operators in (4.53) are the four fermion operators with $d = 6$. As has already been described, there are (up to family indices) only six such operators [4.42,4.43]. They are all of the general form $qqq\ell$, where q and ℓ represent quark and lepton fields, respectively, and family indices have been ignored. These operators can describe nucleon decay and can be generated by the exchange of scalar or vector bosons. For $C \simeq e^2$ they lead to nucleon lifetimes in the interesting range $10^{30} - 10^{33}$ yr for [4.73] M in the range $3 \times 10^{14} - 6 \times 10^{15}$ GeV (c.f. the results for the special case of the SU_5 model). The $qqq\ell$ operators all satisfy $\Delta B = \Delta L$, and those relevant to nucleon decay also satisfy $\Delta S/\Delta B = 0, -1$ and $\Delta I = \frac{1}{2}$ for $\Delta S = 0$ processes. The conservation of B - L by the leading operators does not necessarily imply that B - L is an exact symmetry of the entire theory. In the SU_5 model, the quantum number

$$\begin{aligned} Z &= X + \frac{4}{5} Y \\ X &= \frac{1}{5} N_{10F} - \frac{3}{5} N_{5*F} - \frac{2}{3} N_{5H} - \frac{2}{3} N_{45H} \end{aligned} \tag{4.54}$$

(N_{10F}, N_{5H}, etc., are the number operators for a fermion 10, a Higgs 5, etc.), which coincides with B - L for fermions, is exactly conserved globally by the 't Hooft mechanism [4.69]. In the SO_{10} model, on the other hand, B - L is violated. However, the B - L violation only affects nucleon decay through higher dimension operators (involving ϕ) which represent $X_S - X'$ mixing.

It is possible to write down many other operators with dimension greater than 6 which violate B - L [4.72-4.74]. Because the coefficients of $\mathscr{L}_{eff}$ involve additional inverse powers of M they will only be relevant to nucleon decay if they involve mass scales M smaller than 10^{14-15} GeV. That is, the observation of B - L violating processes would indicate the existence of new thresholds in the desert between M_W and M_X [4.72]. These thresholds could be associated with heavy colored Higgs particles, for example. There is no particular reason to assume the existence of such particles or that their masses should be in an experimentally interesting range (unlike the lepto-quark bosons for which there are independent mass estimates). Nevertheless, it is useful to classify the mass scales relevant to various processes.

The interesting operators of low dimension are listed in Table 4.8. Lorentz, family, and gauge indices are omitted, but can be found in the original papers [4.73-4.74]. For each type of operator Table 4.8 lists the dimension, the relevant ΔB and ΔL, a typical process, and the mass range for which the operator would lead to a nucleon (or, in the $\Delta B = 2$ case, a nuclear) lifetime in the interesting range 10^{30} - 10^{33} yr. These are computed [4.73] by assuming that $C \simeq e^{n-2}$, where n is the number of external fields in the process. This is reasonable if the operator is generated by tree diagrams, provided one counts Yukawa and Higgs quartic couplings as e and e^2, respectively. The entries in Table 4.8 can be understood from a quantity called F-parity, which Weinberg [4.73] has shown must be multiplicatively conserved because of weak SU_2 and Lorentz invariance. F = +1 for q and ℓ and F = -1 for q^c, ℓ^c, $W^{\pm}$, Z, γ, ϕ, $\phi^{\dagger}$, and D, where D represents a derivative. Operators relevant to integer charged quark models are not included in Table 4.8 because they correspond to a low energy theory different from G_S. See Section 3.4.5 and [3.94].

Table 4.8 Interesting operators O of dimension d [4.73-4.74]. q, ℓ, ϕ, and D represent quark, lepton, and Higgs fields, and derivatives, respectively. For the $\Delta B \neq 0$ operators, M is the mass that would lead to a nucleon or nuclear lifetime in the 10^{30} - 10^{33} yr range. For the $\ell\ell\phi\phi$ operator, M is the mass that would imply a neutrino mass in the 10^{-2} - 10^{+2} eV range.

d	O	Process	ΔB, ΔL	M (GeV)
6	$qqq\ell$	$p \to e^+\pi^0$	$\Delta B = \Delta L = -1$	3×10^{14} - 3×10^{15}
7	$qqq\ell^c\phi$	$n \to e^-K^+$	$\Delta B = -\Delta L = -1$	2×10^{10} - 10^{11}
7	$qqq\ell^cD$	$n \to e^-\pi^+$	$\Delta B = -\Delta L = -1$	4×10^{9} - 2×10^{10}
10	$qqq\ell^c\ell^c\ell^c\phi$	$n \to 3\nu$ $n \to \nu\nu e^-\pi^+$	$\Delta B = -\frac{1}{3}\Delta L = -1$	$(3 - 7) \times 10^4$
11	$qqq\ell\ell\ell\phi^2$	$p \to e^+\nu^c\nu^c$	$\Delta B = +\frac{1}{3}\Delta L = -1$	$(2 - 4) \times 10^4$
9	$qqqqqq$	$n \leftrightarrow \bar{n}$ $pn \to \pi^+\pi^0$ $nn \to 2\pi^0$	$\Delta B = -2$ $\Delta L = 0$	4×10^5 - 10^6
5	$\ell\ell\phi\phi$	ν Majorana mass	$\Delta B = 0$ $\Delta L = \pm 2$	10^{11} - 10^{15}

All of the $\Delta B = \pm 1$ operators in Table 4.8 conserve some linear combination of B and L. For example, the d = 7 operators conserve B + L. Wilczek and Zee [4.72] have constructed an example of an SU_5 model to which a 10^9 - 10^{10} GeV Higgs 10 is added. The diagram in Fig. 4.7a generates an effective $qqq\ell^c\phi$

operator (the Yukawa and Higgs couplings of this Higgs 10 necessarily violate B - L). The anti-symmetry of the 10 Yukawa couplings imply [4.72-4.74] that $\Delta S = \bar{1}$ decays are strongly favored by this operator. The other d = 7 operator would dominate in a model with no elementary Higgs fields. M is estimated in this case [4.73] by replacing D by a typical momentum (≃1 GeV).

There has been much interest [4.76-4.81, 4.72] in $\Delta B = \pm 2$ interactions, which could lead to $n \leftrightarrow \bar{n}$ oscillations or nuclear decay through NN annihilation. It is useful to parametrize these effects by a neutron Majorana mass term

$$\mathscr{L}_{eff} = \delta m \; n^T C n + \text{H. C.} \quad , \tag{4.55}$$

which implies $n \leftrightarrow \bar{n}$ oscillations in empty space with a time scale $\tau_{n-\bar{n}} \simeq 1/\delta m$. It could also describe nuclear decay through $nn \to \pi\pi$ or $pn \to \pi\pi$. The relevant time scale is not $\tau_{n-\bar{n}}$, because strong interactions within the nucleus effectively prevent the free $n - \bar{n}$ oscillations from occurring. Instead, one has a nuclear annihilation rate of order [4.76-4.80]

$$\Gamma_{nuc} \simeq a \frac{(\delta m)^2}{m_N} \sim \frac{a}{m_N \tau_{n-\bar{n}}^2} \quad . \tag{4.56}$$

For a nuclear lifetime $\tau_{nuc} = \Gamma_{nuc}^{-1} > 10^{30}$ yr and for $a \simeq 10^{-2}$ (for wave function overlap effects), (4.56) implies $\delta m < 10^{-20}$ eV and $\tau_{n-\bar{n}} > 10^5$ sec. $n - \bar{n}$ oscillation times of order $\gtrsim 10^5$ sec may be observable in reactor experiments. However, the oscillations are greatly suppressed by the earth's magnetic field, which would probably have to be shielded. The phenomenology is further discussed in [4.79], for example.

Marshak and Mohapatra [4.78] have described a model in which $|\Delta B| = 2$ interactions are mediated by the exchange of three Higgs particles Δ_R (which also generate Majorana neutrino masses and give $W_R^\pm$ their masses), as in Fig. 4.7b. This generates

$$\mathscr{L}_{\text{eff}} \simeq \lambda\gamma^3 \frac{<\Delta_R>}{m_{\Delta R}^6} qqq\ qqq \tag{4.57}$$

For $\lambda \simeq \alpha^2$, γ (the Yukawa coupling) $\simeq 10^{-2}$ and $<\Delta_R> \simeq m_{\Delta R} > 10^4$ GeV, one finds a coefficient of $\simeq 10^{-30}$ GeV^{-5} in (4.57), which should correspond to $\delta m \simeq 10^{-21}$ eV. There have also been SU_5 models involving the diagram in Fig. 4.7c, in which two of the Higgs fields are 5's and the third is a 10 [4.72] or 15 [4.81,4.78]. In these cases, however, the effects are enormously suppressed due to the much larger masses of the colored Higgs 5's. A model involving gauge boson exchange and a neutrino Majorana mass term [4.80] also leads to negligibly small effects.

Feinberg, Goldhaber, and Steigman [4.82] have discussed the $\Delta B = \Delta L = \pm 2$ process $pp \to e^+e^+$, which could lead to nuclear annihilation or (in empty space) $H \leftrightarrow \bar{H}$ oscillations. They argue that the effective four fermi interaction strength must be less than 10^{-21} G_F from the absence of gamma rays from $H\bar{H}$ annihilation in interstellar space, and $<10^{-20}$ G_F from nuclear stability. The Marshak-Mohapatra model [4.78], with the fourth Higgs line attached to leptons, leads to an effective interaction eighteen orders of magnitude smaller than these limits.

Weinberg [4.73] has discussed the cosmological implications of new interactions (see also [4.72]). He argues that a baryon asymmetry produced by the "ordinary" B - L conserving interactions (Chapter 5) would be washed out at later times if there are additional baryon number violating interactions, unless these new interactions conserve some combination B + aL. This argues against the existence of $\Delta B = 2$, $\Delta L = 0$ interactions or against the existence of more than one new interaction.