PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

From the Standard Model to Grand Unification

J. Ellis

Phil. Trans. R. Soc. Lond. A 1982 **304**, 69-86
doi: 10.1098/rsta.1982.0006

Phil. Trans. R. Soc. Lond. A **304**, 69–86 (1982) [69]
Printed in Great Britain

From the standard model to grand unification

By J. Ellis
*CERN, Theory Division, CH-*1211 *Geneva* 23, *Switzerland*

This paper reviews the limitations of the standard SU(3) × SU(2) × U(1) model and develops the philosophy of grand unification. Some simple grand unified theories are presented, and calculations made of the order of magnitude of the fine-structure constant α, as well as of $\sin^2 \theta_W$ and some quark masses. Predictions for nucleon decay and neutrino masses are then discussed; they may be observable in the near future. It is suggested that grand unified theories complex enough for the understanding of the baryon asymmetry of the Universe may also predict a neutron electric dipole moment large enough to be measured. Finally, some inadequacies of GUTs are mentioned.

1. Introduction

This paper contains an assessment of the limitations of the standard SU(3) × SU(2) × U(1) gauge model of the fundamental interactions, and proceeds to introduce grand unified theories (GUTs) embracing the strong, weak and electromagnetic interactions (Ellis 1980). Some implications of GUTs for particle physics will be described, but cosmological applications are largely left to the paper by Steigman (this meeting). At the end of this paper various unsolved problems will be mentioned whose resolution must await a still grander synthesis.

The first part of the paper emphasizes the inadequacies of the standard SU(3) × SU(2) × U(1) model (Glashow 1961, Salam 1968, Weinberg 1967) and mentions some empirical regularities of the unexplained parameters of the fundamental fermion mass spectrum which point to a possible philosophy (Georgi & Glashow 1974) for developing further the unification of the fundamental interactions. Then, after a discussion of the general philosophy of grand unification – which leads already to quite restrictive bounds on the magnitude of the fine-structure constant α (Ellis & Nanopoulos 1981) – some simple GUTs based on the groups SU(5) and SO(10) will be presented as examples. In §6 various predictions will be obtained from GUTs for previously undetermined parameters of the standard model, such as the neutral weak mixing angle θ_W (Georgi *et al.* 1974) and the masses of several quarks, including most notably that of the bottom quark (Chanowitz *et al.* 1977). Then predictions for new types of fundamental interactions will be developed, most notably for those violating baryon number conservation – which should lead to nucleon decays with a detectable lifetime – and for violations of lepton number which may generate masses for the neutrinos. There follow some brief remarks about the necessity to go beyond minimal GUTs if one is to explain the baryon asymmetry observed in the Universe (Sakharov 1967, Ellis *et al.* 1980*d*), and the ensuing likelihood that the neutron electric dipole moment is large enough to be detected in forthcoming experiments (Ellis *et al.* 1981*a*, *b*). Finally, the most glaring inadequacies of GUTs will be highlighted.

2. Limitations of the standard model

The preceding papers of this meeting have shown that we have a satisfactory understanding of QCD and the Glashow–Salam–Weinberg SU(2) × U(1) weak interaction theory, at least in their perturbation theoretic aspects. We have also seen much experimental confirmation that these theories give a satisfactory description of particle interactions in the range of energies presently accessible. Therefore, the standard SU(3) × SU(2) × U(1) model is probably a correct low energy approximation to the full dynamics of the fundamental interactions. However, there

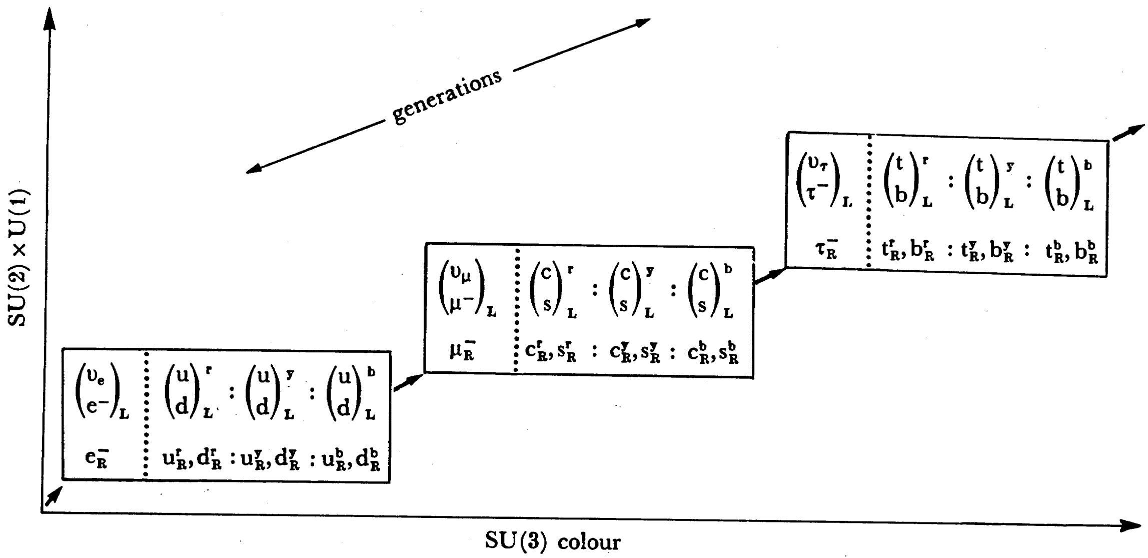

Figure 1. Schematic picture of the generation structure of fundamental fermions.

are clearly many aspects in which the standard model is deficient. The theory is not completely unified, as it has three different gauge-coupling constants $g_3 \neq g_2 \neq g_1$, and it has no explanations or predictions for many of the fundamental observable quantities. Thus it has no explanation for charge quantization ($|Q_e|/|Q_p| = 1 + O(10^{-20})$) and offers no understanding of the quark and lepton masses or of the weak mixing angles. Indeed the standard model contains at least 20 arbitrary parameters (27 or more if one allows for neutrino masses). They are: three gauge-coupling constants g_3, g_2 and g_1, two non-perturbative θ-vacuum parameters, six quark masses and four assorted Kobayashi–Maskawa mixing parameters, two parameters to characterize the Higgs potential, and three lepton masses (six lepton masses and four mixing parameters if neutrinos have masses).

In attempting to unify further the fundamental interactions, we seek some empirical indications of the direction in which to proceed, and a strong hint is provided by the apparent 'family' or 'generation' pattern of the fundamental fermions. This structure is portrayed in figure 1. The known fermions seem to occur in sets of 15 helicity states with masses that are qualitatively similar. Thus the lightest 'generation' comprises the u and d quarks, the electron and its neutrino, while the second contains (c, s, μ, ν_μ) and the third contains (t, b, τ, ν_τ). The known fundamental forces act mainly on fermions in the same generation. This is exactly true for strong and electromagnetic interactions at zero momentum transfer, and true to a very good approximation for the neutral weak interactions mediated by the Z^0. It is less true

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

for the charged weak interactions mediated by the W±, but even so the generalized Cabibbo–Kobayashi–Maskawa (1973) mixing angles are apparently quite small.

A cynic might question whether this generation structure is merely a figment of an overly credulous gestalt mechanism. However, Froggatt & Nielsen (1979) have made a Monte-Carlo analysis that suggests very strongly that the generation pattern is real, and also motivates the philosophy of grand unification that embodies it. They analyse whether the perceived pattern of fermion masses is in fact random, or whether there are statistically significant correlations. To this end they assume the existence of a set of Abelian charges which they assign randomly to left- and right-handed fermion fields f_L and f_R. They then assume that the Dirac mass terms $m\bar{f}_R f_L$ are given by the following simple *ansatz* in terms of the Abelian charges:

$$m = M_0 \exp(-\nu|Q_L - Q_R|) \tag{1}$$

where M_0 is an overall scale parameter, ν is a number, and $|Q_L - Q_R|$ is some reasonable norm in the space of Abelian charges. They then compute the statistical distribution of the ratios of masses of fermions of the same charge, (e.g. d, s, b; u, c, t), in the forms of the quantities $\ln(m_3/m_2)/\ln(m_2/m_1)$. They also compute a covariance matrix

$$\tilde{C}_{ij} \equiv \frac{(\ln m_i - \langle \ln m_i \rangle)(\ln m_j - \langle \ln m_j \rangle)}{\langle \ln(m_2/m_1)\rangle^2}, \tag{2}$$

which measures the correlation in mass between fermions of different charges, as well as values for the generalized Cabibbo–Kobayashi–Maskawa mixing angles. When they compare the predictions with the experimental values of these quantities they find that for fermions of the same charge

$$0.6 \lesssim \ln(m_3/m_2)/\ln(m_2/m_1) \lesssim 0.7 \tag{3}$$

experimentally, whereas in their model they find that $\ln(m_3/m_2)/\ln(m_2/m_1) \approx 0.6 \pm 0.2$. Thus their 'random' model is consistent with the pattern of masses of fermions of the same charge. However, when they compute the covariance matrix $\tilde{C}_{ij}$ (equation (2)), they find that the model's matrix elements are typically about six times the experimental values. This strongly suggests that the masses of the fermions of different charges are indeed 'correlated' as suggested by the generation picture of figure 1. Moreover, they find that, even if quarks of different charges have similar masses, in their model there is no tendency for the weak mixing angles to be small. They conclude that there are statistically significant correlations between the fermions of different charges insofar as they exhibit similar masses and small mixing, which are 'suggestive of a non-Abelian unification of both left- and right-handed fermions'.

This is indeed the direction pursued in GUTs. These theories strive first to unify the interactions between the different members of each generation of quarks and leptons. Since these are put in common multiplets, there will in general be direct quark–lepton transitions (see also Pati & Salam 1973*a*, *b*, 1974). Ultimately one may hope to go further and explain the number of different generations as well as the hierachical ratios of their masses. However, a fully satisfactory scheme of this type has yet to emerge, and this paper will concentrate on the grand unification of interactions within each generation.

3. The philosophy of grand unification

We shall seek to embed the interactions of the standard model into a simple group G:

$$G \rightarrow SU(3) \times SU(2) \times U(1) \rightarrow SU(3) \times U(1). \tag{4}$$

We may hope thereby to obtain explanations of some of the old mysteries, notably charge quantization and predictions for exciting new quark–lepton interactions that may cause nucleons to decay, and a reduction in the total number of parameters in the theory, for example by determining the ratios of quark and lepton masses.

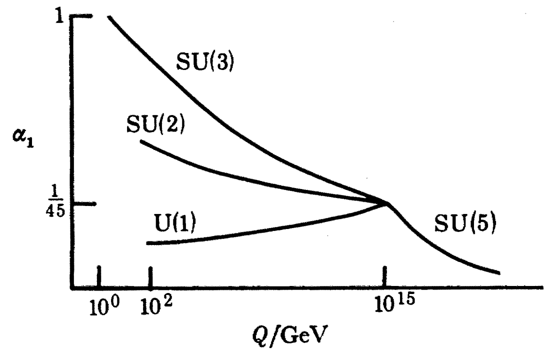

Figure 2. Illustration of the way in which the SU(3), SU(2) and U(1) coupling constants approach each other, as a function of energy, in a GUT such as SU(5).

There is a major hurdle to be jumped in this programme, namely the unification of gauge couplings. We know that experiments at present energies find that

$$g_3 \gg g_2 \neq g_1. \tag{5}$$

However, we also know from the renormalization group (Stueckelberg & Peterman 1953, Gell-Mann & Low 1954) that field-theory couplings vary with the energy Q at which they are evaluated, and in particular that g_3 decreases because of asymptotic freedom (Politzer 1973, Gross & Wilczek 1973):

$$g_3(Q^2) \approx 12\pi/[(33-2N_q) \ln (Q^2/\Lambda^2)] \tag{6}$$

where N_q is the number of quarks with masses much less than Q, and Λ is an arbitrary scale parameter. The other couplings g_2 and g_1 also depend on energy, but less strongly than g_3 unless there are many quarks. We therefore have in mind a picture like that in figure 2, where all three couplings may come together at some higher energy scale. Because of the logarithmic evolution (6) of the coupling constants, this grand unification scale is generally very large. Indeed, if we take Λ of order 1 GeV and $\alpha = 1/137$ to characterize the non-strong interactions, we find a grand unification scale of order 10^{15} GeV (Georgi *et al.* 1974, Buras *et al.* 1978, Ellis *et al.* 1980*d*). While enormous, this scale is still much less than 10^{19} GeV, the Planck energy scale at which quantum gravity effects become of order unity: because the Newton constant $G_N \approx 10^{-38}$ GeV^{-2} it follows that $G_N Q^2 \approx 1$ at $Q \approx 10^{19}$ GeV.

It may be instructive to observe that the fine-structure constant must in fact lie within quite restrictive bounds if the grand unification philosophy is to make any sense (Glashow & Nanopoulos 1979, Ellis & Nanopoulos 1981). The reason is that grand unification cannot take place at a scale less than 10^{14} GeV, otherwise the nucleon lifetime would be less than the present experimental limit of about 10^{30} years. Also, grand unification cannot be postponed beyond 10^{19} GeV, otherwise it makes no sense not to include gravitation in the unification programme. In a leading approximation to the renormalization group equations we have

$$\frac{1}{\alpha} = 9+\frac{11}{\pi}\ln\frac{m_X}{m_W}+\frac{8}{3}\frac{1}{\alpha_3(m_W)}, \tag{7}$$

where 9 is the renormalization of the effective $\alpha(Q)$ between zero momentum transfer, the Thompson limit, and $Q \approx m_W$. Assuming that $\alpha_3(Q) \approx 1$ when $Q \approx 1$ GeV so that $0.1 \lesssim \alpha_3(m_W) \lesssim 0.2$, we find from equation (7) that

$$\tfrac{1}{120} \lesssim \alpha \lesssim \tfrac{1}{170}. \tag{8}$$

It seems too good to be a coincidence that the experimental value of α lies in this range, and it seems churlish to spurn this opportunity to attempt grand unification. Accordingly, let us study some simple models.

4. Simple models for grand unification

It has been shown (Georgi & Glashow 1974) that the only group of rank four suitable for grand unification is SU(5). Let us see what it gives. The theory contains 24 gauge vector bosons, 12 of which are the familiar γ, $W^{\pm}$, Z^0 and $g_1, \ldots, g_8$. In addition, there are 12 gauge bosons that must be very massive if nucleons are not to decay too fast: the coloured weak doublet ($X_{R,Y,B}$, $Y_{R,Y,B}$) and their antiparticles. In addition, there are at least $15N_G$ fermion helicity states, where N_G is the number of generations. It is convenient to use left-handed two-component fields to represent all these fermions; thus we use $\bar{q}_L$ instead of q_R, etc. The 15 helicity states of each generation are placed into a reducible $\bar{\mathbf{5}}+\mathbf{10}$-dimensional representation, taking the following form for the first generation of figure 1:

$$\bar{\mathbf{5}} = \begin{bmatrix} \bar{d}_R \\ \bar{d}_Y \\ \bar{d}_B \\ e^- \\ \nu_e \end{bmatrix}_L, \quad \mathbf{10} = \sqrt{\tfrac{1}{2}} \begin{bmatrix} 0 & \bar{u}_B & -\bar{u}_Y & -u_R & -d_R \\ -\bar{u}_B & 0 & \bar{u}_R & -u_Y & -d_Y \\ \bar{u}_Y & -\bar{u}_R & 0 & -u_B & -d_B \\ u_R & u_Y & u_B & 0 & -e^+ \\ d_R & d_Y & d_B & e^+ & 0 \end{bmatrix}_L. \tag{9}$$

The strong SU(3) subgroup acts on the first three indices of SU(5), while the weak SU(2) subgroup acts on the last two indices. It is then simple to read off from expressions (9) their SU(3) × SU(2) decompositions:

$$\bar{\mathbf{5}} = (\bar{3}, 1)+(1, 2), \quad \mathbf{10} = (3, 2)+(\bar{3}, 1)+(1, 1). \tag{10}$$

The two $(\bar{3}, 1)$ representations of SU(3) × SU(2) are just right for the left-handed $\bar{u}$ and $\bar{d}$ since they are colour antitriplets and weak singlets. We identify the $(\bar{3}, 1)$ in the $\bar{\mathbf{5}}$ as the $d_{R,Y,B}$ as this fixes charge quantization correctly:

$$3Q_{\bar{d}}+Q_{e^-} = 0 \Rightarrow Q_d = -\tfrac{1}{3}. \tag{11}$$

The unit charge of the W^+ then fixes $Q_u = +\frac{2}{3}$ and therefore $Q_{\text{proton}} = 2Q_u+Q_d = +1$ as desired. The (3, 2) in the **10** is just right for the $(u, d)_{R,Y,B}$ coloured weak doublet, while the leptons are easily assigned to the colour singlets in equation (10). Assigning heavier generations to identical representations, we find that the SU(5) model reproduces all the conventional strong, weak and electromagnetic interactions of the fundamental fermions. As an aside we note that the SU(5) anomalies of the $\bar{\mathbf{5}}$ and **10** representations are equal and opposite, so that the theory is renormalizable.

To break the symmetry in the desired fashion (4) we need both an adjoint **24** of Higgs for the first stage and a fundamental **5** of Higgs for the second stage:

$$\mathrm{SU}(5) \xrightarrow[\mathbf{24}\text{ of Higgs}]{} \mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1) \xrightarrow[\mathbf{5}\text{ of Higgs}]{} \mathrm{SU}(3)\times\mathrm{U}(1). \qquad (12)$$

The **24** of Higgs can be written as a traceless 5×5 matrix ϕ whose vacuum expectation value can be written in real and diagonal form as

$$\langle 0|\phi|0\rangle = O(10^{15})\times\begin{bmatrix} 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & -\frac{3}{2} & 0\\ 0 & 0 & 0 & 0 & -\frac{3}{2}\end{bmatrix}\mathrm{GeV}. \qquad (13)$$

This form obviously leaves invariant the SU(3) of the first three indices and the SU(2) of the last two indices, as well as a U(1) relative phase transformation. It will also generate masses of order 10^{15} GeV for the X and Y bosons coupling together the colour and weak SU(2) subgroups. It is apparent from the representation contents (9) that the exchanges of X and Y bosons mediate interactions that violate baryon number B and lepton number L conservation, while conserving the combination B–L:

$$\mathrm{u}+\mathrm{u}\rightarrow\bar{\mathrm{d}}+\mathrm{e}^+,\quad \mathrm{u}+\mathrm{d}\rightarrow\bar{\mathrm{u}}+\mathrm{e}^+,\quad \mathrm{u}+\mathrm{d}\rightarrow\bar{\mathrm{d}}+\nu_\mathrm{e}. \qquad (14)$$

We shall return to more details of these interactions. The **5**-dimensional representation of Higgs fields H is called upon to violate weak SU(2) and hence has a vacuum expectation value of the form

$$\langle 0|H|0\rangle = O(10^{2})\times\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\mathrm{GeV} \qquad (15)$$

giving masses to the fermions, $W^\pm$ and Z^0, in the usual way. It also leads to generalized Cabibbo–Kobayashi–Maskawa (Kobayashi & Maskawa 1973) mixing between the different generations, as we shall see later.

The next simplest GUT is based on the rank-five group SO(10) (Georgi 1974, Fritzsch & Minkowski 1975, Chanowitz *et al.* 1977). There are several possible patterns of symmetry-breaking, including

$$\mathrm{SO}(10)\begin{array}{l}\nearrow \mathrm{SU}(5)\rightarrow\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)\searrow\\ \searrow \mathrm{SU}(4)\times\mathrm{SU}(2)\times\mathrm{SU}(2)\rightarrow\ldots\nearrow\end{array}\mathrm{SU}(3)\times\mathrm{U}(1). \qquad (16)$$

We shall not dwell on this theory here, but we note that it contains 45 gauge vector bosons, including additional superheavy bosons capable of mediating nucleon decay, beyond those already found in the SU(5) subgroup. Each generation is now represented by an irreducible

16-dimensional representation, related to the SU(5) representations (9) by the decomposition

$$\mathbf{16} = \mathbf{10} + \bar{\mathbf{5}} + \mathbf{1}, \tag{17}$$

where the singlet has charge equal to zero and its colourless, and so can be identified with a left-handed antineutrino field (or right-handed neutrino). Many different representations of Higgs fields have been proposed for roles in the spontaneous breakdown of SO(10) to SU(3) × U(1), including **10**, **16**, **45**, **54**, **120** and **126**-dimensional representations (Ellis 1980). The most economical pattern of symmetry breaking is

$$\mathrm{SO}(10) \xrightarrow{\mathbf{16}} \mathrm{SU}(5) \xrightarrow{\mathbf{45}} \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \xrightarrow{\mathbf{10}} \mathrm{SU}(3) \times \mathrm{U}(1), \tag{18}$$

where the **45** contains the previous **24** of SU(5), and the **10** is a $\mathbf{5} + \bar{\mathbf{5}}$ of SU(5). Charge quantization works in the same way as in the SU(5) model, and freedom from anomalies is automatic because SO(10) is a 'safe' group. One possible advantage of this model over SU(5) is that each generation is now described by an irreducible representation, but this advantage is somewhat vitiated by the necessity to make at least three copies of it.

Many larger groups have been proposed for GUTs, including E_6 and SU(8), but most of the generic features are present in the two examples given above. Accordingly we shall use them to illustrate the general features of GUTs developed in the rest of this paper.

5. Determination of the grand unification scale

The general principles of fixing the grand unification scale were given in §3. The SU(3), SU(2) and U(1) couplings are different at present energies and approach each other at a rate given by the renormalization group, becoming equal at energy scales greater than or approximately equal to $m_{X,Y}$ of the very massive gauge bosons. For example, the SU(3) and SU(2) interactions approach each other at a rate

$$\frac{1}{\alpha_3(Q)} - \frac{1}{\alpha_2(Q)} \approx -\frac{11}{12\pi} \ln\,(m_X^2/Q^2), \tag{19}$$

in the leading logarithmic approximation. To make a more refined estimate of m_X we must include various higher-order effects (Ross 1978, Marciano 1979, Goldman & Ross 1979, Ellis *et al.* 1980*d*) such as two-loop terms in the renormalization group equations which reduce the estimate of m_X by one quarter, threshold effects at energies of about m_W and m_X which reduce m_X by about a further one sixth, and a low-mass Higgs boson which reduces the estimate of m_X by one half. One must also be careful to take into account the variation of the fine structure constant α between $Q = 0$ and $Q = m_W$ which was mentioned in §3. It is mainly due to the simple fermion loops of figure 3 and has the dramatic effect of reducing the estimate of m_X by one tenth. When all these effects are taken into account we find that the principal uncertainty in m_X is that due to our lack of knowledge of the QCD scale parameter Λ. In terms of the modified minimal subtraction scheme at the two-loop order with four flavours, we find (Ellis *et al.* 1980*d*)

$$10^{15}\Lambda_{\overline{\mathrm{m.s.}}} \lesssim m_{X,Y} \lesssim 2 \times 10^{15}\Lambda_{\overline{\mathrm{m.s.}}}, \tag{20}$$

if there is no substantial new physics in the desert region between now and 10^{15} GeV, and we take a minimal GUT such as SU(5) in which there is no intervening stage of symmetry-breaking before the X boson is reached.

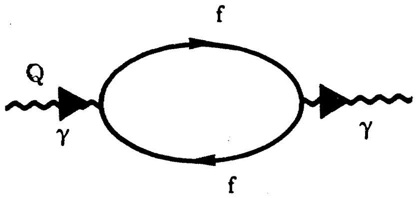

FIGURE 3. Dominant contribution due to fermion loops to the renormalization of α between $Q = 0$ and $Q \approx 100$ GeV.

6. Predictions for low-energy parameters

In view of the enormous scale of grand unification symmetry-breaking, we might wonder whether it is possible to make any useful predictions based on the symmetry. The answer is yes, because the renormalization group enables us to calculate symmetry-breaking effects (Georgi *et al.* 1974).

Take for example the parameter $\sin^2\theta_W$ related to the ratio of g_2 and g_1 in the weak interaction theory. Models such as SU(5) and SO(10) containing conventional fermion generations predict a symmetry value

$$\sin^2\theta_W(Q \gtrsim m_X) = \tfrac{3}{8}. \tag{21}$$

However, $\sin^2\theta_W$ is a function of energy scale that is renormalized at lower energies:

$$\sin^2\theta_W(Q) = \tfrac{3}{5}g_1^2(Q)/[g_2^2(Q)+\tfrac{3}{5}g_1^2(Q)], \tag{22}$$

and its variation can be computed by using the renormalization group equations for g_1 and g_2. One finds that at lower energies

$$\sin^2\theta_W(Q) \approx \frac{3}{8}\left[1-\frac{\alpha}{4\pi}\left(\frac{110-N_H}{9}\right)\ln\left(\frac{m_X^2}{Q^2}\right)\right]+O(\alpha) \tag{23}$$

where N_H is the number of light Higgs doublets, usually taken to be one. Formula (23) gives $\sin^2\theta_W \approx 0.2$ when it is evaluated (Buras *et al.* 1978) on an energy scale around 100 GeV. Knowing the effective value of $\sin^2\theta_W$ relevant to the experiments done at present energies requires the computation of many radiative corrections. The final result (Marciano & Sirlin 1981) is

$$\sin^2\theta_W^{\text{eff}} = 0.210+0.004(N_H-1)+0.006\ln(0.4/\Lambda_{\overline{\text{m.s.}}}) \tag{24}$$

at an average Q^2 of 20 GeV^2. To fit low energy neutral current cross sections requires a second parameter in the Glashow–Salam–Weinberg model, namely the strength parameter ρ defined by

$$\mathscr{L}^{\text{eff}} = \sqrt{\tfrac{1}{2}}G_F[J_\mu^+J^{-\mu}+\rho J_\mu^0 J^{0\mu}]. \tag{25}$$

The parameter ρ takes the value one in the tree approximation if SU(2) is only broken by $I = \frac{1}{2}$ Higgs fields, but is subject to radiative corrections that alter it (Marciano & Sirlin 1981) to between 0.99 and 1.00 in the minimal SU(5) model. Figure 4 shows a comparison between ρ and $\sin^2\theta_W$ extracted from experimental data (Liede & Roos 1981 private communication). Despite the tremendous precision of the experimental data, they disagree with the GUT prediction by only about one standard deviation. Perhaps GUTs are on the right track?

Another set of low energy predictions of GUTs concern quark and lepton masses. Since GUTs place quarks and leptons in the same representation of a large group, one can expect

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

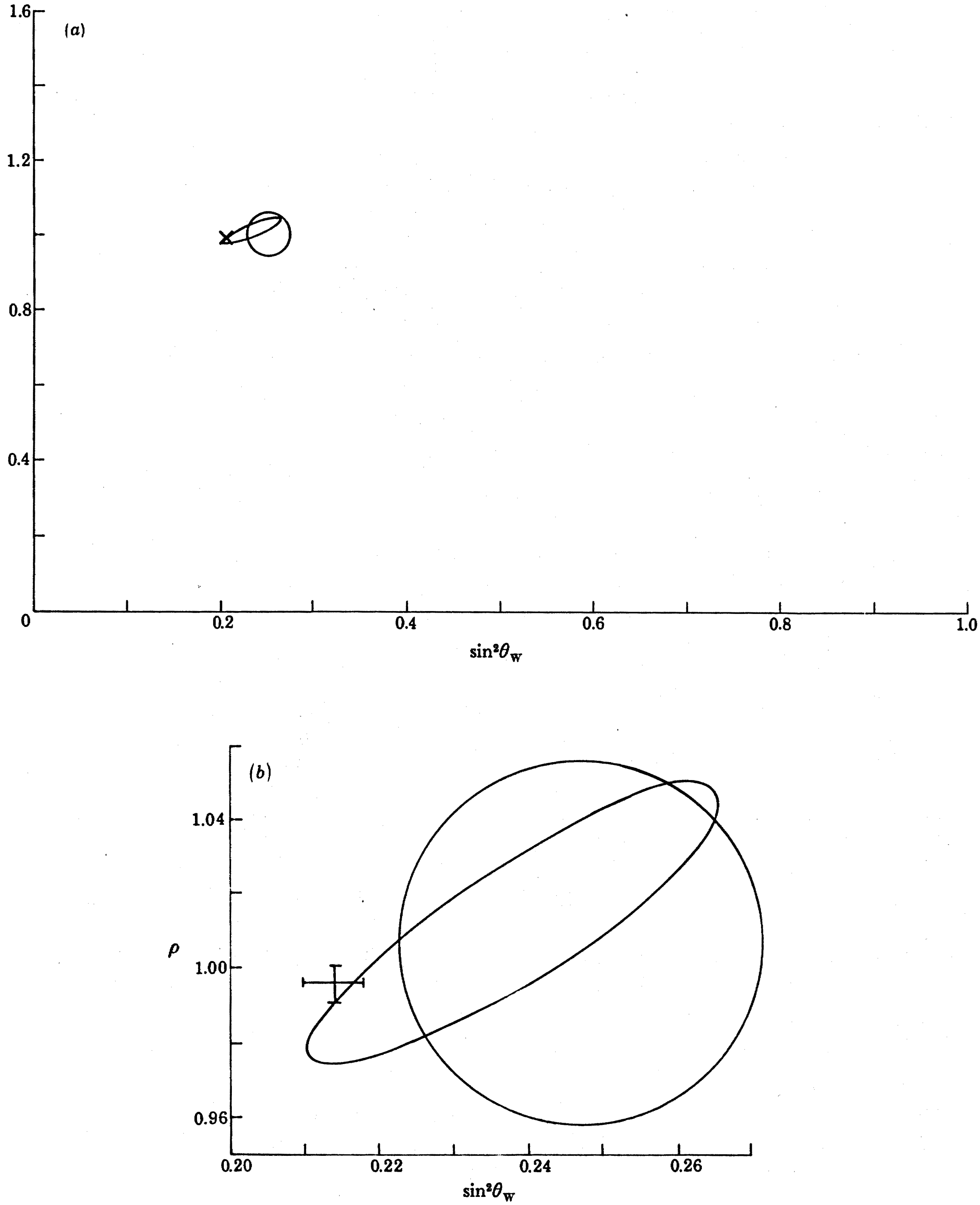

FIGURE 4. (*a*) Experimental determination of the neutral current parameters ρ and $\sin^2 \theta_W$. (*b*) Detailed comparison with the predictions of the SU(5) GUT. Closed loops, fits to data; crosses, SU(5) GUT prediction.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

symmetry predictions for m_q/m_l ratios in many theories (Chanowitz *et al.* 1977, Buras *et al.* 1978). For example, in SU(5) and economical versions of SO(10) we have the relations

$$m_e/m_d = m_\mu/m_s = m_\tau/m_b = 1 \tag{26}$$

in the symmetry limit, corresponding to an SU(4) symmetry of $\langle 0|H|0\rangle$ in equation (15). The predictions (26) get renormalized in a calculable way at low energies when SU(5) is broken. We define effective scale-dependent masses $m_f(Q)$ from the inverse fermion propagators

$$S_f^{-1}(Q) = \not{Q} - m_f(Q). \tag{27}$$

The predictions (26) apply to the effective masses at scales $Q \gtrsim m_X$. The dominant renormalizations of these effective masses come from the diagrams of figure 5, which can be summed by using the renormalization group. One obtains for example

$$[m_b(Q)/m_\tau(Q)] = [\alpha_3(Q)/\alpha_5(m_X)]^{4/(11-\frac{2}{3}N_q)}\,[1+O(20\%)\text{ corrections}]^{-1} \tag{28}$$

where corrections come from SU(2), U(1) and higher-order SU(3) corrections to the leading-order gluon exchange. Note the dependence of the ratio (28) on the total number of quarks N_q. When we calculate the effective mass at the e^+e^- threshold for producing $b\bar{b}$ quark pairs, $Q = 2m_b(Q)$, we find

$$5 \lesssim m_b \lesssim 5\tfrac{1}{2}\text{ GeV} \tag{29}$$

if there are only six quarks in total. This prediction for m_b is increased unacceptably if there are eight or more quarks (Nanopoulos & Ross 1979). A similar argument based on equation (26) leads to the qualitative prediction that

$$m_s \approx \tfrac{1}{2}\text{ GeV}. \tag{30}$$

There is a considerable discussion as to whether this prediction is quantitatively correct (Weinberg 1977, Ellis 1980), but it is at least in the right parish. One question on which there is no disagreement is the fact that the renormalization-invariant prediction

$$m_s/m_d = m_\mu/m_e \tag{31}$$

is incorrect: the left-hand side is generally believed to be $O(20)$, while the right-hand side is about 200. Nevertheless, m_d is at least predicted to be small, and this qualitative success together with the predictions (29) and (30) constitute the first solid indication that the naïve generation assignments of figure 1 are actually correct, and bear out the statistical arguments (Froggatt *et al.* 1979) of §2. Some GUTs go on to make predictions for the top quark mass: one interesting possibility raised (Barbieri & Nanopoulos 1980, Glashow 1980) in considerations of E_6 and other GUTs is that

$$m_t = m_c(m_\tau/m_\mu) \Rightarrow m_t \approx 20\text{ GeV}, \tag{32}$$

a prediction that can soon be tested at PETRA.

After this survey of GUT predictions for parameters of the standard $SU(3)\times SU(2)\times U(1)$ model, we now consider GUT predictions for exciting new interactions.

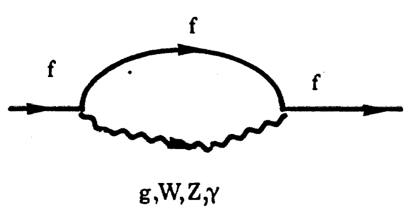

FIGURE 5. Dominant contribution to the renormalization of the quark : lepton mass ratios.

7. Nucleon decay

Today very few theorists believe that baryon-number conservation is the sacred principle it used to seem (Stueckelberg 1939, Wigner 1949). The underlying theoretical reason is that since the advent of gauge theories it has become apparent that the only exact symmetries of nature are gauge symmetries. Examples are colour SU(3) and electromagnetic U(1), both of which are conventional exact gauge symmetries associated with massless spin-1 bosons, the gluon and the photon. Another example is Lorentz invariance, with general relativity interpretable as a gauge theory of the Lorentz group with the massless graviton as the corresponding gauge boson. Any massless gauge boson coupled to baryon number (or lepton number) must have a coupling constant $g \lesssim O(10^{-20})$ to be compatible with its non-observation. It seems likely that no such massless gauge boson exists, and therefore that one may expect baryon-number conservation to be violated in some way. Theoretical mechanisms for baryon non-conservation do in fact come about via non-perturbative effects in the Weinberg–Salam weak interaction theory ('t Hooft 1976*a*, *b*) and in general relativity (Zeldovich 1976, Hawking *et al.* 1979). In the former case weak instantons cause $\Delta B = N_G$ transitions, where N_G is the number of fermion generations, which would of course not cause protons or bound neutrons to decay. In general relativity, it has been known for some time that the only conserved quantum numbers that a black hole can 'remember' are electromagnetic charge, colour, mass and spin. This means that we can imagine using a black hole as a 'catalyst' for a baryon-number violating proton (p) $\rightarrow$ positron (e^+) transition

$$\mathrm{p} + (\text{black hole}) \rightarrow (\text{black hole})' \rightarrow \mathrm{e}^+ + (\text{black hole}). \tag{33}$$

In fact it has been suggested that proton decay could be mediated by virtual mini-black-holes with a mass of the order of the Planck mass, *ca.* 10^{19} GeV, which gives a lifetime

$$10^{45} \lesssim \tau_{\text{proton}} \approx m^4_{\text{Planck}}/m^5_{\text{proton}} \lesssim 10^{50} \text{ years.} \tag{34}$$

It is unfortunate that the detection of such a long lifetime is beyond the reach of currently imaginable experiments.

It is fortunate that GUTs predict much faster nucleon decays through the effective four-fermion interaction mediated by the exchanges of superheavy bosons. For example, in SU(5), if one neglects generalized Cabibbo–Kobayashi–Maskawa mixing one has (Buras *et al.* 1978) the following effective interaction mediated by X and Y exchanges:

$$\mathscr{L}_{\mathrm{GU}} = 2\sqrt{2}\, G_{\mathrm{GU}} [(\epsilon_{ijk} \bar{\mathrm{u}}^{\mathrm{c}}_{k\mathrm{L}} \gamma_\mu \mathrm{u}_{j\mathrm{L}}) (2\bar{\mathrm{e}}^+_{\mathrm{L}} \gamma^\mu \mathrm{d}_{i\mathrm{L}} - \bar{\mathrm{e}}^+_{\mathrm{R}} \gamma^\mu \mathrm{d}_{i\mathrm{R}}) + (\epsilon_{ijk} \bar{\mathrm{u}}^{\mathrm{c}}_{k\mathrm{L}} \gamma_\mu \mathrm{d}_{j\mathrm{L}}) (\bar{\nu}^{\mathrm{c}}_{l\mathrm{R}} \gamma^\mu \mathrm{d}_{i\mathrm{R}}) + \text{Hermitian conjugate}] \tag{35}$$

where we can write

$$g^2/8m^2_{\mathrm{X}} \approx g^2/8m^2_{\mathrm{Y}} \equiv \sqrt{\tfrac{1}{2}}\, G_{\mathrm{GU}} \tag{36}$$

by analogy with the corresponding expression for the Fermi weak interaction,

$$g^2/8m^2_{\mathrm{W}} = \sqrt{\tfrac{1}{2}}\, G_{\mathrm{F}}. \tag{37}$$

Let us focus on some qualitative features of the baryon decays mediated by (35). Clearly the decay amplitude $A \propto 1/m^2_{\mathrm{X}}$ and hence the decay rate

$$\Gamma \propto 1/m^4_{\mathrm{X}}. \tag{38}$$

Since it has the dimensions of mass, the decay rate Γ (38) must be scaled by some mass parameter to the fifth power, and the most likely one is the nucleon mass m_N:

$$\Gamma_N = C^{-1} m_N^5 / m_X^4 \Rightarrow \tau_N = C m_X^4 / m_N^5. \tag{39}$$

If we suppose that $C = O(1)$ in the generic formula (39), and if we set $m_X = O(10^{15})$ GeV as discussed earlier, then we find that the nucleon lifetime should be about 10^{30} years. This is expected to be much smaller than the lifetime induced by mini-black-holes (34) simply because m_X is expected to be much smaller than the Planck mass.

If we want to make a more precise prediction for the nucleon lifetime and decay modes, then it is clear from equation (39) that in addition to using the more precise estimate (20) of m_X we need to calculate the decay rate factor C, which will in general depend on the precise form of $\mathscr{L}_{GU}$ and on the bound-state dynamics of the nucleon. We now address these issues more closely. An important question is how the different generations should appear mixed in the true forms of the interaction (35). What are the generalized GUT Cabibbo–Kobayashi–Maskawa angles? This question cannot be given a general answer, but one does find that in a reasonably general class of GUTs the GUT angles are very closely related to the familiar weak angles. The interested reader is referred elsewhere for the proof (Ellis *et al.* 1980*b*, *d*); here is just quoted the corrected Cabibbo-rotated form of (35) for two generations:

$$\begin{aligned}\mathscr{L}_{GU} = 2\sqrt{2}\, G_{GU}\, e^{i\phi_X} [\![(\epsilon_{ijk} \bar{u}^c_{kL} \gamma_\mu u_{jL} \{ [(1+\cos^2\theta_C)\, \bar{e}^+_L + \sin\theta_C \cos\theta_C \bar{\mu}^+_L]\, \gamma^\mu d_{iL} \\ + [(1+\sin^2\theta_C)\, \mu^+_L + \sin\theta_C \cos\theta_C\, \bar{e}^+_L]\, \gamma^\mu s_{iL} + (\bar{e}^+_R \gamma^\mu d_{iR}) + (\bar{\mu}^+_R \gamma^\mu d_{iR}) \} \\ + (\epsilon_{ijk} \bar{u}^c_{kL} \gamma_\mu)\, [d_{iL} \cos\theta_C + s_{jL} \sin\theta_C]\, [\bar{\nu}^c_{eR} \gamma^\mu d_{iR} + \nu^c_{\mu R} \gamma^\mu s_{jR}] + \text{Hermitian conjugate}]\!]. \end{aligned} \tag{40}$$

which reduces to (35) when we set $\theta_C \to 0$ and restrict to the first generation.

Accepting the effective interaction (40), what selection rules can we deduce? Clearly we always have

$$\Delta S/\Delta B \lesssim 0 \Rightarrow p, n \not\longrightarrow K^- + X, \tag{41}$$

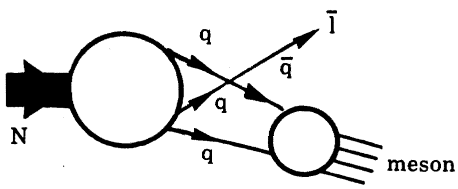

FIGURE 6. Dominant mechanism for nucleon decay in GUTs.

which could in fact have been deduced already from the (qqql) form of the interaction. Moreover, in addition to observing that most baryon decays should be 'Cabibbo-favoured', we can also make numerical predictions for the Cabibbo-disfavoured decays:

$$\frac{N \to \mu^+ + \text{non-strange}}{N \to e^+ + \text{non-strange}} \approx \frac{\sin^2\theta_C \cos^2\theta_C}{(1+\cos^2\theta_C)^2 + 1}, \tag{42a}$$

$$\frac{N \to e^+ + \text{strange}}{N \to \mu^+ + \text{strange}} \approx \frac{\sin^2\theta_C \cos^2\theta_C}{(1+\sin^2\theta_C)^2 + 1}. \tag{42b}$$

It will be very important to test the form of the effective low energy interaction (40) and test the predictions (42) *if* baryon decay is ever detected. This is because baryon decay may be our only available experimental window on 10^{15} GeV physics, and these predictions go to the guts of our GUTs, probing the soft underbelly of the ill-understood generation structure and Higgs sector. The qualitative successes of the mass predictions $m_d = m_e$, $m_s = m_\mu$, $m_b = m_\tau$ give some circumstantial support for the usual naïve generation structure associating u, d with e; c, s with μ; and t, b with τ; but more direct evidence for this hypothesis would be helpful.

Armed with our estimate (20) of m_X and (40) of $\mathscr{L}_{GU}$ we can now calculate the nucleon lifetime and branching ratios. The calculations proceed in two stages analogous to those for non-leptonic weak decays. The effective interaction (40) is a 'bare' one corresponding to a very-short-range interaction taking place at a distance $\Delta x \approx 1/m_X \approx 10^{-28}$ cm. To calculate its matrix elements between conventional hadrons of size 10^{-13} cm we must 'dress' it with all strong, weak and electromagnetic interactions that can take place on scales between 10^{-28} and 10^{-13} cm. These correspond mainly to the exchanges of $SU(3) \times SU(2) \times U(1)$ gauge bosons with momenta between m_N and m_X. They result in an enhancement of the decay amplitude by a factor (Ellis *et al.* 1980*d*, Wilczek & Zee 1979) A in the range $\frac{7}{2} \lesssim A \lesssim 4$, decreasing the baryon lifetime by a factor $O(15)$ compared with what we would have got without 'dressing' the 'bare' operator (40).

We should now estimate the matrix elements of the 'dressed' operator using conventional ideas from hadron phenomenology. It is generally believed that the dominant mechanism for baryon decay is that illustrated in figure 6 in which two of the nucleon's quarks annihilate to give an antilepton, and an antiquark that combines with the spectator quark to form mesons. The problem then reduces to calculating $\langle M|\mathscr{L}_{GU}|N\rangle$ in some suitable approximation – a difficult problem in hadron physics. Two extreme approaches have been taken: they treat the quarks involved with completely non-relativistic kinematics and naïve SU(6) spinology such as

$$|p\uparrow\rangle = \sqrt{\tfrac{1}{2}}\,u\uparrow(u\uparrow d\downarrow - u\downarrow d\uparrow), \tag{43}$$

or an ultra-relativistic bag model. While the numerical estimates are qualitatively similar, even specific calculations with the same method can produce different results (Ellis *et al.* 1980*d*):

$$0.6(m_X/5\times 10^{4}\ \text{GeV})^4 \lesssim \tau_{p,n} \lesssim 25(m_X/5\times 10^{14}\ \text{GeV})^4. \tag{44}$$

If we take a central value of $m_X = 6\times 10^{14}$ GeV, corresponding to $\Lambda_{\overline{m.s.}} = 400$ MeV in formula (20) for minimal SU(5), and allow an uncertainty of $2^{\pm 1}$ for higher-order effects, uncertainties in $\Lambda_{\overline{m.s.}}$, etc., we eventually obtain

$$\tau_{p,u} \approx 8\times 10^{30}(\times 10^{0\pm 2})\ \text{years}. \tag{45}$$

It is unfortunately impossible to be more precise than this, even in the minimal SU(5) model, and the uncertainties are even larger in non-minimal GUTs such as SO(10).

Decay branching ratios into individual mesonic final states can be estimated by using similar non-relativistic or ultra-relativistic assumptions. A comparison between the results so obtained, together with the results of a 'preferred' middle-of-the-road model of Kane & Karl (1980), are shown in table 1. We see that although there are some variations between the branching ratios under different assumptions, there is a reasonable consensus on the following being substantial decay modes of the proton and neutron respectively:

$$p \to e^+\omega,\ e^+\pi^0; \quad n \to e^+\pi^-. \tag{46}$$

While experiments should be (and are) designed to be sensitive to a broad range of different decay modes, it is encouraging that the two-body $e^+\pi$ decay modes look so promising, as these

TABLE 1. NUCLEON DECAY BRANCHING RATIOS IN MINIMAL SU(5)
(Extracted from Kane & Karl (1980).)

	decay mode	non-relativistic model	preferred 'recoil' model	relativistic model
p	e^{+0}	21	25	26
	$e^+\rho^0$	2	7	11
	$e^+\pi^0$	36	40	38
	$e^+\eta$	7	2	0
	$\nu\rho^+$	1	3	4
	$\nu\pi^+$	14	16	15
	μ^+K^0	18	8	5
	$\nu_\mu K^+$	0	0	1
n	$\nu\omega$	5	5	5
	$\nu\rho^0$	1	1	2
	$\nu\pi^0$	8	7	7
	$\nu\eta$	2	0	0
	$e^+\rho^-$	6	12	19
	$e^+\pi^-$	79	72	68
	$\nu_\mu K^0$	1	3	1

have clean experimental signatures. The experimental problems of detecting baryon decay are discussed in Fiorini (1981, this symposium). We theorists wish our experimental colleagues luck with the predictions of equation (45) and table 1.

8. NEUTRINO MASSES

We now turn to lepton-number-violating interactions in GUTs, and their possible implications for neutrino masses. We know that neutrino masses are much smaller than those of their companion leptons:

$$m_{\nu_e}/m_e < 10^{-4}, \quad m_{\nu_\mu}/m_\mu < \tfrac{1}{200}, \quad m_{\nu_\tau}/m_\tau < 0.14, \tag{47}$$

but there is no fundamental reason why the m_ν should be zero. Indeed the gauge theory dogma that the only exact gauge symmetries are gauge symmetries, and the absence of a massless boson coupled to lepton number L, lead one to expect a breakdown of L-conservation, which can in general lead to neutrino masses. In contrast to the masses of conventional quarks and charged leptons which come from the 'Dirac' couplings of Higgs fields to left- and right-handed components,

$$(H_{I=\frac{1}{2}})\,\overline{f}_R f_L, \tag{48}$$

neutrino masses can also come from a 'Majorana' interaction of the type

$$(H_{I=1,\Delta l=2})\,\nu_L\,\nu_L. \tag{49}$$

This can occur because the neutrino has zero electric charge and colour, so that the antineutrino (a right-handed field) can usurp the role of a right-handed neutrino field. An interaction of the type (49) is not possible in the minimal Glashow–Salam–Weinberg model which only has $I = \frac{1}{2}$ Higgs fields. Neither does it occur in the minimal SU(5) GUT which contains no $I = 1$ Higgs field coupled to fermion pairs. This model conserves B–L, and since a neutrino mass term has $B = 0$, this also implies L-conservation in this case. However, neutrino masses occur naturally in more complicated GUTs that contain more Higgs fields or fermions, or both (Ellis 1980). In general, we may encounter interactions of the type (49), with

$$\langle 0|H_{I=1}|0\rangle = O(m_W^2/m_X) \tag{50}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

(Magg & Wetterich 1980, Barbieri *et al.* 1980), and hence

$$m_\nu \approx (m_t \text{ or } m_W)^2/m_X, \tag{51}$$

being a generic feature of such GUTs. Alternatively, one may encounter an effective low energy interaction of the form

$$M^{-1}[(H_{I=\frac{1}{2}})\,\nu_L]\,[(H_{I=\frac{1}{2}})\,\nu_L], \tag{52}$$

which could be due to heavy particle exchange. One generally expects such superheavy particles to have masses $M = O(10^{15\pm4})$ GeV, so that the general estimate (51) applies to this type of mass term also. One then sees that to the extent that $m_W/m_X \ll 1$, so also is $m_\nu/m_t \ll 1$.

Another possibility present in many GUTs such as SO(10) (Georgi 1975, Fritzsch *et al.* 1975, Chanowitz *et al.* 1977) is that there exists an SU(3) × SU(2) × U(1) singlet right-handed neutrino field ν_L^c (or $\bar{\nu}_R$). In this case one could expect a 'Dirac' coupling of the type (48) and therefore a mass term of the general order of magnitude of a conventional fermion (quark or lepton) mass. But in this case one can also have a 'Majorana' mass for the ν_R, and one should rather consider a mass matrix problem:

$$(\nu_L, \bar{\nu}_R)\begin{bmatrix} \approx (m_t \text{ or } m_W)^2/M & \approx (m_t \text{ or } m_W) \\ \approx (m_t \text{ or } m_W) & \approx M \end{bmatrix}\begin{bmatrix} \nu_L \\ \bar{\nu}_R \end{bmatrix}, \tag{53}$$

where the top left-hand matrix element is taken from equation (51), and the off-diagonal elements from equation (48). The $\bar{\nu}_R\bar{\nu}_R$ 'Majorana' mass term is generally expected to be of order 10^{15} GeV or so, since the ν_R is an SU(3) × SU(2) × U(1) singlet state. Diagonalizing the mass matrix (53) we again find a light neutrino mass eigenstate with a mass of the order (51). If we put $m_t \approx 30$ GeV (corresponding to the t quark?) and $M = 10^{15}$ GeV then we find $m_\nu \approx 10^{-3}$ eV. But in some GUTs one easily discovers that M can be smaller than m_X. For example, in the minimal version of SO(10) one finds (Witten 1980) that

$$M \approx O(\alpha/\pi)^2 m_X \tag{54}$$

and in this case the neutrino mass could be as high as 10 eV. This may be a plausible order of magnitude for the upper bound on the neutrino mass in GUTs. On the other hand M could be as high as 10^{19} GeV, the scale of quantum gravity, in which case $(\langle 0|H|0\rangle)^2/M$ gives a neutrino mass of *ca.* 10^{-5} eV. A plausible range for neutrino masses in GUTs may therefore be

$$O(10^{-5})\text{ eV} \lesssim m_\nu \lesssim O(10)\text{ eV}, \tag{55}$$

and it will be very interesting to see some experimental light cast on this interesting domain. This may be done either by direct mass measurements (Lubimov *et al.* 1980, De Rújula 1981) or indirectly through neutrino oscillations (Maki *et al.* 1962, Pontecorvo 1957, 1958, 1967).

9. Cosmology and the neutron electric dipole moment

Although the cosmological implications of GUTs are not the main topic of this paper, there is an interesting suggestion from cosmology for another low-energy observable consequence of GUTs (Ellis *et al.* 1981 *a*, *b*). It can be argued that GUTs complicated enough to explain the baryon asymmetry in the Universe will in general also predict a neutron electric dipole moment not much smaller than the present experimental upper limit. The CP-violating neutron electric dipole moment d_n may arise from two different sources: conventional weak interaction perturbation theory and the non-perturbative QCD vacuum parameter θ. It has been estimated (Baluni 1979, Crewther *et al.* 1979) that the θ-contribution to d_n is

$$\Delta d_n \approx 4\times 10^{-16}\,\theta \text{ e cm}, \tag{56}$$

and from the present experimental upper limit of 6×10^{-25} e cm (Altarev *et al.* 1981) we could deduce that $\theta < 1.5\times 10^{-9}$. We have no idea why θ should be so small, and we can isolate various contributions to θ that are not zero in general. The expression (56) is in terms of an effective θ measured on a momentum scale of an order 1 GeV. Perhaps there is a theory of everything (TOE) that predicts the effective θ measured on some very large momentum scale (Ellis & Gaillard 1979). Between this large scale and 1 GeV, we must traverse the GUT threshold and the weak interaction threshold, and these types of non-strong interactions may renormalize θ:

$$\theta(1\text{ GeV}) = \theta_{\text{TOE}} + \delta\theta_{\text{GUT}} + \delta\theta_{\text{KM}}. \tag{57}$$

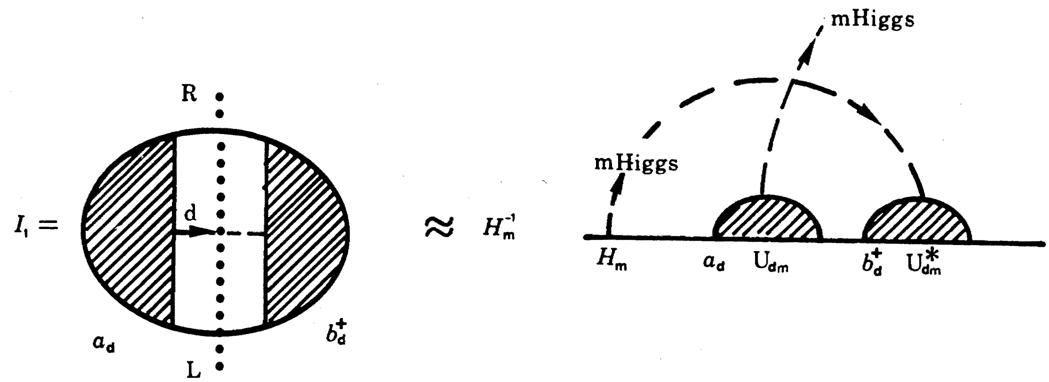

FIGURE 7. Relation between the CP-violating asymmetry in heavy Higgs particle (d) decay and a contribution to $\delta\theta_{\text{GUT}}$. The mHiggs is the field responsible for the masses of the fermions.

The weak renormalization $\delta\theta_{\text{KM}}$ has been estimated (Ellis *et al.* 1979) as $O(10^{-16})$, and the resulting contribution (56) to d_n is actually smaller than the conventional weak interaction perturbation theory contribution. It seems that in many GUTs there are diagrams contributing to $\delta\theta_{\text{GUT}}$ that have a very similar coupling structure to those responsible for the CP-violating asymmetry in the decays of heavy Higgs particles which are presumed to be responsible for generating the baryon asymmetry of the Universe, as illustrated in figure 7. This connection enables us to derive a qualitative lower bound on d_n in terms of the baryon:photon ratio (n_{B}/n_γ):

$$d_n \gtrsim 2.5\times 10^{-18}(n_{\text{B}}/n_\gamma) \text{ e cm}. \tag{58}$$

If one takes the cosmological estimate (Steigman 1981) that $(n_{\text{B}}/n_\gamma) \gtrsim 2\times 10^{-10}$, one finds that d_n should be as large as 5×10^{-28} e cm, about three orders of magnitude smaller than the present experimental upper limit. A new series of experiments is under way to improve this limit by a few orders of magnitude: perhaps a neutron electric dipole moment will be found soon?

10. Inadequacies of grand unified theories

While they are more elegant than the standard model, and make many striking predictions for quantities observable at present energies, some of which agree well with experiment, our present GUTs are clearly inadequate in many respects. Let us recall some of them.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Clearly GUTs do not predict everything, indeed the minimal SU(5) GUT contains at least 23 parameters (30 if one allows for neutrino masses): one gauge coupling g_5 and one non-perturbative vacuum parameter θ_5, nine parameters for the Higgs potential, six quark and lepton masses (three more for neutrinos) and six generalized Cabibbo–Kobayashi–Maskawa mixing parameters (four more if neutrinos have masses and mix). Obviously we would like to find a more fundamental theory with fewer parameters.

So far we have no explanation or prediction for the number of generations. There is a phenomenological preference for keeping to three generations, based on the b quark mass prediction and supported by cosmological considerations on neutrino counting, but no fundamental understanding of the number of generations. Indeed, we have no theory of fermion masses, or of their weak mixing, which are left as unclear as before we started grand unification.

There is the fundamental problem (Gildener 1976, Buras *et al.* 1978) of the hierarchy of mass scales in GUTs: Why and how is $m_W/m_X \ll 1$? At our present level of understanding this seems to require an unnatural adjustment of parameters in the Higgs potential.

Cosmology may present some problems, in that minimal GUTs predict too few baryons (Ellis *et al.* 1980*c*) and perhaps too many monopoles (Preskill 1979). However, the first of these difficulties is cured in slightly non-minimal GUTs, and there may be no need to complicate the GUT to suppress the primordial production of monopoles (Bais & Rudaz 1980).

Perhaps some of the problems listed above will be solved if we try more seriously (Ellis *et al.* 1980*a*, *e*) to include gravity in our unification of particle interactions, but this is the subject of another paper at this symposium (Salam 1981).

References (Ellis)

Altarev, I. S. *et al.* 1981 Leningrad Institute of Nuclear Physics preprint 636.
Bais, F. A. & Rudaz, S. 1980 *Nucl. Phys.* B **170** (FS1), 507.
Baluni, V. 1979 *Phys. Rev.* D **19**, 2277.
Barbieri, R. & Nanopoulos, D. V. 1980 *Physics Lett.* B **91**, 369.
Buras, A. J., Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. 1978 *Nucl. Phys.* B **135**, 66.
Chanowitz, M. S., Ellis, J. & Gaillard, M. K. 1977 *Nucl. Phys.* B **128**, 506.
Crewther, R. J., Di Vecchia, P., Veneziano, G. & Witten, E. 1979 *Physics Lett.* B **88**, 123.
De Rújula A. 1981 CERN preprint TH.3045.
Ellis, J. 1980 CERN preprint TH.2942.
Ellis, J. & Gaillard, M. K. 1979 *Nucl. Phys.* B **150**, 141.
Ellis, J. Gaillard, M. K. Maiani, L. & Zumino, B. 1980*a* In *Unification of the fundamental particle interactions* (ed. S. Ferrara, J. Ellis & P. Van Nieuwenhuizen). New York: Plenum Press.
Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. 1980*b* *Physics Lett.* B **88**, 320.
Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. 1980*c* In *Unification of the fundamental particle interactions* (ed. S. Ferrara, J. Ellis & P. Van Nieuwenhuizen). New York: Plenum Press.
Ellis, J., Gaillard, M. K., Nanopoulos, D. V. & Rudaz, S. 1980*d* *Nucl. Phys.* B **176**, 61.
Ellis, J., Gaillard, M. K., Nanopoulos, D. V. & Rudaz, S. 1981*a* *Physics Lett.* B **99**, 101.
Ellis, J., Gaillard, M. K., Nanopoulos, D. V. & Rudaz, S. 1981*b* *Nature, Lond.* **293**, 41.
Ellis, J., Gaillard, M. K. & Zumino, B. 1980*e* *Physics Lett.* B **94**, 343.
Ellis, J. & Nanopoulos, D. V. 1981 *Nature, Lond.* **292**, 436.
Fritzsch, H. & Minkowski, P. 1975 *Ann. Phys.* **93**, 193.
Froggatt, C. D. & Nielsen H. B. 1979 *Nucl. Phys.* B **164**, 114.
Gell-Mann M. & Low F. E. 1954 *Phys. Rev.* **95**, 1300.
Georgi, H. & Glashow S. L. 1974 *Phys. Rev. Lett.* **32**, 438.
Georgi, H., Quinn H. R. & Weinberg, S. 1974 *Phys. Rev. Lett.* **33**, 451.
Gildener, E. 1976 *Phys. Rev.* D **14**, 1667.
Glashow, S. L. 1961 *Nucl. Phys.* B, **22**, 579.
Glashow, S. L. 1980 *Phys. Rev. Lett.* **45**, 1914.
Glashow, S. L. & Nanopoulos, D. V. 1979 *Nature, Lond.* **281**, 464.

Goldman, T. & Ross, D. A. 1979 *Physics Lett.* B **84**, 208.
Gross, D. J. & Wilczek, F. A. 1973 *Phys. Rev. Lett.* **30**, 1343.
Hawking, S., Page, D. N. & Pope, C. N. 1979 *Physics Lett.* B **86**, 175.
Kane, G. & Karl, G. 1980 *Phys. Rev.* D **22**, 2808.
Kobayashi, M. & Maskawa, K. 1973 *Prog. theor. Phys.* **49**, 652.
Lubimov, V. A. *et al.* 1980 *Physics Lett.* B **94**, 61.
Magg, M. & Wetterich, C. 1980 *Physics Lett.* B **94**, 61.
Maki, Z., Nakagawa, M. & Sakata, S. 1962 *Prog. theor. Phys.* **28**, 870.
Marciano, W. J. 1979 *Phys. Rev.* D **20**, 274.
Marciano, W. J. & Sirlin, A. 1981 *Phys. Rev. Lett.* **46**, 163.
Pati, J. C. & Salam, A. 1973*a* *Phys. Rev. Lett.* **31**, 661.
Pati, J. C. & Salam, A. 1973*b* *Phys. Rev.* D **8**, 1240.
Pati, J. C. & Salam, A. 1974 *Phys. Rev.* D **10**, 275.
Politzer, H. D. 1973 *Phys. Rev. Lett.* **30**, 1346.
Pontecorvo, B. 1957 *Soviet Phys. JETP* **33**, 549.
Pontecorvo, B. 1957 *Soviet Phys. JETP* **34**, 247.
Pontecorvo, B. 1967 *Soviet Phys. JETP* **53**, 1717.
Ross, D. A. 1978 *Nucl. Phys.* B **140**, 1.
Sakharov, A. D. 1967 *Pis'ma Zh. éksp. teor. Fiz.* **5**, 32.
Salam, A. 1967 In *Proc. 8th Nobel Symposium* (ed. N. Svartholm). Stockholm: Almqvist and Wiksells.
Stueckelberg, E. C. G. 1939 *Helv. phys. Acta* **11**, 299.
Stueckelberg, E. C. G. & Peterman, A. 1953 *Helv. phys. Acta* **26**, 499.
't Hooft, G. 1976*a* *Phys. Rev. Lett.* **37**, 8.
't Hooft, G. 1976*b* *Phys. Rev.* D **19**, 3432.
Weinberg, S. 1967 *Phys. Rev. Lett.* **19**, 1264.
Weinberg, S. 1977 *I. I. Rabi Festschrift, Trans. N.Y. Acad. Sci.* II, 38.
Wigner, E. P. 1949 *Proc. Am. phil. Soc.* **93**, 521.
Wilczek, F. A. & Zee, A. 1979 *Phys. Rev. Lett.* **43**, 1571.
Witten, E. 1980 *Physics Lett.* B **91**, 81.
Zeldovich, Ya. B. 1976 *Physics Lett.* A **59**, 254.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES