Unified Field Theories*

Hitoshi Murayama

Department of Physics, University of California

Berkeley, CA 94720

and

Theoretical Physics Group

Lawrence Berkeley National Laboratory

University of California

Berkeley, CA 94720

---

*Appeared in *Encyclopedia of Applied Physics*, Vol. **23**, 1 (1998)

Abstract

This article explains the idea of unified field theories in particle physics. It starts with a historical review of two successful theories which unified two apparently distinct forces: Maxwell's theory of electromagnetism and Glashow–Salam–Weinberg theory of electroweak forces. Then it describes known four forces in Nature: gravitational, electromagnetic, weak and strong forces, and what their similarities and differences are. All of them are described based on the same principle (gauge principle), despite their distinct appearances due to their different phases of dynamics. Especially the last three forces are very similar to each other theoretically. Moreover, they seem to be intimately related to each other in order to guarantee the neutrality of matter (why the hydrogen atom is absolutely charge neutral) and seemingly miraculous anomaly cancellation in the electroweak theory which otherwise renders the quantum theory inconsistent.

Grand Unified Theories (GUTs) are introduced to describe the last three forces as consequences of a single force, and explain the neutrality of matter and anomaly cancellation automatically. The original model is disfavored both from experimental and theoretical point of view. Experimentally, it predicts a lifetime of proton below experimental lower bound, and the pattern of measured gauge coupling constants does not follow its prediction. Theoretically, it predicts the energy scale of unification many orders of magnitudes larger than the scale of the weak force and makes the electroweak theory unstable against the radiative corrections. Supersymmetry is introduced to solve the latter problem. It makes the hierarchy in energy scales

stable against the radiative corrections, in a similar fashion that the discovery of the positron made the electron rest energy stable against its own Coulomb interaction in electrodynamics. It turns out that the incorporation of supersymmetry to GUTs changes the predicted pattern of the gauge coupling constants which beautifully agrees with measurements.

Kaluza–Klein theories attempt unification of gravity with other three forces, by seeking for origin of the latter forces in gravity of higher-dimensional space time, *e.g.* 11 dimensions. Even though the idea is beautiful, no realistic models could be constructed so far. It is difficult to obtain a parity-violating (chiral) matter content as we see experimentally, and also has a problem of ultraviolet divergences beyond our control.

The current hope of unifying gravity with three other forces is in superstring theory. It replaces elementary particles by small strings with a typical size of 10^{-33} cm. The theory can incorporate gravity without ultraviolet divergences, and allows chiral matter content with realistic gauge groups. Furthermore, the theory contains only one coupling constant and hence, if true, explains all parameters in particle physics by the single constant. One major problem in string theory is that it allows too many solutions to their equations of motion within perturbation theory and hence is not predictive. Non-perturbative effects are hoped to eliminate many of the perturbative solutions and solve the problem. Recent progress in understanding no-perturbative dynamics of string theory is briefly reviewed. It still lacks concrete and testable phenomenological predictions, but there is a constant activity towards its goal.

Glossary

Unified Theories	Theories of elementary particles and forces among them which attempt a unified description of many forces as different manifestations of a single force.
Supersymmetry	A hypothesized symmetry which interchanges bosons and fermions, whose spins differ by 1/2. If supersymmetry exists, every particle in the standard model of particle physics, such as quarks, leptons, gauge bosons and Higgs bosons must have their superpartners.
anomaly	Suppose a theory defined at the level of classical Lagrangian has a symmetry. It sometimes happens that the symmetry is broken once the theory is quantized. This phenomenon in quantum field theory is called anomaly. For instance, the scale invariance of QCD Lagrangian is broken quantum mechanically and the coupling constant runs as a function of energy, *i.e.*, the theory is no longer scale invariant. This anomaly is needed to explain the asymptotic freedom of strong interaction and does not lead to any inconsistencies. On the other hand, the gauge invariance of a theory may also be broken quantum mechanically in chiral theories, which makes the theory inconsistent; it cannot maintain renormalizability and unitarity simultaneously. Therefore, one has to check that all gauge anomalies are canceled in chiral theories. This requirement puts a stringent constraint on the matter content of the theory.
chiral	The word means handedness in Greek. A theory is called chiral if

the right-handed and left-handed helicity states of a single particle interact differently with gauge fields at relativistic energies. A chiral theory necessarily breaks parity. The Glashow–Weinberg–Salam theory of electroweak forces is chiral.

compactification Suppose the true dimensionality of space-time is not four but is larger, such as 10 in superstring theory. One assumes that extra dimensions ($10-4=6$ in this case) have very small spatial volume so that they are not visible in the current experiments. In this case, the extra dimensions are said to be compactified. The compactification can occur as a solution to the equations of motion of Einstein's equation with matter generalized to higher dimensional space-time.

Contents

Introduction

The idea of unification is to look for a single theory from which all observed phenomena can be ultimately derived. Even though the idea appears unrealistically optimistic at first sight, it has a good grounding in the context of elementary particle physics. In fact, history has shown us examples that many varied phenomena could be understood in terms of a single and simple theory. This article intends to explain historic examples, attempts for finding unified description of different phenomena, and the most recent activities.

1 Prototype Unification

The Maxwell's theory of electromagnetism has been regarded as one of the best example of unification. Another good example is the unified theory of electromagnetism and weak interactions. In both cases, there were early experimental hints of a need for a unified description, theoretical progress towards such a direction, prediction of new phenomena based on unified theories, and experimental confirmation of the unified theories.

By the end of the 19th century, it was clear that the Faraday's notion of the field is the most useful description of both electric and magnetic forces. This was probably the only theoretical similarities between these two apparently distinct forces. On the other hand, there were experimental indications that the two forces were related. An electric current creates a magnetic field around it. A change in the magnetic flux induces an electric potential.

Maxwell pushed it to the extreme. Since a change in the magnetic flux can induce an electric field, they might be two manifestation of a single force. If it is true, a change in the electric field must also induce the magnetic field. Even though there were no experimental indication of it by then, Maxwell postulated the existence of such a new phenomenon, which allowed him to formulate the theory of electric and magnetic fields on equal footing. In particular, such a unified description predicted the existence of an electromagnetic wave, which consists of electric and magnetic fields without any sources propagating in the vacuum by inducing each other. Finally the existence of electromagnetic wave was confirmed and the Maxwell's theory was established.

Furthermore, Maxwell's theory of electromagnetism lead to a further revolution in physics. The Maxwell's equations happen to have a symmetry which was first recognized by Lorentz and Poincaré. The symmetry, known as Lorentz invariance, was later emphasized as a true symmetry of space-time by Einstein in his special theory of relativity, which required us to change our basic notion of absolute time.

The Glashow–Weinberg–Salam theory of electromagnetic and "weak" forces is another example that provided a unified description of two different forces by a single theory which further predicted a new force, the neutral-current weak interaction, and a new particle, the Higgs boson. We will discuss the electroweak theory further in the next section, after briefly describing the known four forces in Nature.

Table 1: Comparison of four known forces in Nature. The existence of the graviton is not confirmed yet. See the text for the meanings of "Universality" and "Gauge invariance."

	Gravity	Electromagnetism	Strong	Weak
Range	∞	∞	10^{-13} cm	10^{-16} cm
Force Carrier	graviton (?)	photon	gluon	W, Z bosons
(spin)	(2)	(1)	(1)	(1)
Gauge invariance	general coordinate tr.	U(1)	SU(3)	SU(2)
Universality	equivalence principle	electric charge	SU(3) rep.	SU(2) rep.
source	anything with energy	charged particles	hadrons	left-handed quarks, leptons
theory	general relativity	Maxwell's theory	QCD	electroweak theory
	classical	classical/quantum	quantum	quantum
phase	Coulomb(?)	Coulomb	confinement	Higgs
coupling constant	Newton's constant	fine-structure constant	strong coupling constant	weak coupling constant

2 Four Forces

There are at least four known basic forces in Nature: gravity, electromagnetism, "weak" and "strong" forces. Here, weak and strong refer to the proper names of two kinds of forces which reflect their strengths as well.

The four forces manifest themselves in quite distinctive ways. Below are their brief descriptions. See Table 1.

2.1 Gravity

The gravitational force was first formulated in Newton's theory of gravity. It has an infinite range with a potential $V(r) \propto 1/r$, and acts on any bodies with a mass. Its remarkable feature is that the motions induced by gravity do not depend on the composition or mass of the body. Any body follows the same trajectory. This is called the equivalence principle. Because of this property, Einstein proposed his theory of general relativity, which describes the gravity as a curvature of four-dimensional space-time. According to this theory, moving bodies follow geodesics: they appear to fall because the space-time is curved. This is a natural explanation why gravity exhibits the equivalence principle. The quantum theory of gravity is still not available; in such a quantum theory, gravity is supposed to be mediated by a spin-two carrier, the graviton. There is a single coupling constant in the theory, Newton's constant $G_N = (M_{\text{Planck}})^{-2} = (1.2 \times 10^{19}\ \text{GeV})^{-2}$, which has a dimension of inverse mass squared in the natural unit $\hbar = c = 1$.

2.2 Electromagnetism

The electromagnetic force also has an infinite range with a Coulomb potential $V(r) \propto 1/r$. It acts on any matter with electric charges. Like gravity, the form of the interaction is completely determined only by the charge of an object, irrespective of its composition. This universality is now understood based on the gauge principle, which allows us to change the phase of quantum mechanical wave functions or quantum fields locally. The coupling constant of the theory is the fine-structure constant, $\alpha = e^2/4\pi\epsilon_0\hbar c \simeq 1/137$, which is a dimensionless number. With this only one parameter, electromagnetism and especially its quantum version (Quantum ElectroDynamics, or QED) is extraordinarily successful, and is in agreement with experimental data at the precision of 10^{-11} (Kinoshita 1995). Due to quantum effects, the fine-structure constant actually varies slowly with energy scales, growing for larger energies.

2.3 Strong Force

The strong force is an interaction among the hadrons, such as baryons (proton, neutron, hyperons, *etc.*) and mesons (π, K, ρ, *etc.*), and is responsible for various phenomena, which include nuclear α-decay, fission, fusion, nuclear binding force and hadron resonances. It has a finite range $\simeq 10^{-13}$ cm and appeared completely different from above two forces. Now the hadrons are understood as composites made up of quarks: baryons as bound states of three quarks, and mesons of a quark and an anti-quark. The quarks come in three kinds, called "colors," and one can rotate three kinds of colors arbitrarily (SU(3) gauge invariance) analogous to the phase rotation in the QED. The color quantum number is the source of the gluon field, just like the electric charges for the electric field. The theory is characterized by a dimensionless coupling constant, the strong coupling constant α_s. The gluon mediates a long-range force, but with a much stronger variation with energy scales than the fine-structure constant. Experimentally, $\alpha_s(91 \text{ GeV}) = 0.118 \pm 0.003$, which is much stronger than the fine-structure constant. The strong coupling constant is smaller at higher energies (asymptotic freedom), grows logarithmically at lower energies, and becomes practically infinite around 300 MeV. Because of this behavior of the coupling constant, the potential among colored particles is approximately $V(r) \propto r$ at large distance $r > 10^{-13}$ cm, and the quarks are "confined" in bound states, *i.e.*, it requires an infinite amount of energy to separate a quark out from a bound state.

The observed strong interactions among hadrons are understood as residual force among the bound states similar to the van der Waals force (residual Coulomb force) among the atoms (electron-nucleus composites). Such a strong residual force exhibited highly non-universal (composition and mass dependent) and non-perturbative phenomena which hindered the formulation of theory until the 1970's. The experimental discovery that there is an approximate scaling in the deep inelastic electron-nucleon scattering experiments at SLAC in the late 1960's by the SLAC-MIT group (Bloom *et al.*, 1969; Breidenbach *et al.*, 1969), and a theoretical discovery that a non-abelian gauge theory can exhibit asymptotic freedom by Gross and Wilczek (1973), Politzer (1973) enabled people to formulate Quantum ChromoDynamics (QCD) based on SU(3) gauge invariance. Because QCD becomes perturbative at high energies, many quantitative tests of QCD were and are being performed at high-energy experiments successfully.

2.4 Weak Force

The weak force manifested itself in nuclear β-decay first, and later observed in various meson decays, muon decays and neutrino interactions. It has the smallest range among the forces observed, $\simeq 10^{-16}$ cm. Fermi formulated the first theory of weak force which showed a remarkable universality. The β-decay lifetimes of many different nuclei and mesons could be described by a single Fermi constant $G_F = 1.1664 \times 10^{-5}$ GeV^{-2}. In particular, the strengths of the weak force acting on the muon and up-quark are measured accurately and agree at the level of a permille. Glashow (1961) speculated that the weak force could also be a gauge force based on $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ group because of this universality. An important point here is that the β-decay changes the electric charge of the parent nuclei ("charged-current weak interaction"). If the weak force is mediated by a gauge boson (named the W-boson) similar to the photon or gluon, it must carry an electric charge. The only way a gauge boson (W) can interact with another gauge boson (photon) is when they belong to the same non-abelian gauge group (SU(2) in this case). Therefore a gauge theory of the weak interaction needs also incorporate the electromagnetic force on equal footing: a need for a unified description. To be consistent with the observed pattern of electric charges, Glashow needed another group, $\mathrm{U}(1)_Y$, and predicted a new kind of weak force: the neutral-current weak interaction mediated by the Z-boson. Such a new type of weak force was discovered later by neutrino experiments, first by the Gargamelle Collaboration (Hasert *et al.*, 1973), and is now tested at permille accuracies at LEP (an e^+e^- storage ring at CERN, Geneva) and SLC (a linear e^+e^- collider at Stanford Linear Accelerator Center) (for a review, see P. Langacker in Particle Data Group, 1996). One feature of the weak force which turns out to be crucial in building unified models is that it distinguishes the left-handed from right-handed helicity states of particles. Thus the weak force is said to be *chiral*, as opposed to the other three forces which do not distinguish the helicities of particles of particles: *vector-like*.

A puzzle in Glashow's theory is that why the weak force mediated by W- or Z-boson is short-ranged, while the theory also contains the photon.

The explanation to this question is the Higgs mechanism. In the Meißner effect in a superconductor, a condensation of charges (Cooper pairs) results in a finite range of a gauge force, namely the finite penetration length of the magnetic field into a superconductor. In a similar fashion, a condensation of yet-to-be-found Higgs boson causes a finite range of the weak force. This theory was formulated by Weinberg (1967) and Salam (1968), based on a spontaneous breakdown of SU(2)×U(1) gauge group to the electromagnetic U(1) gauge group by the Higgs mechanism at the energy scale of 250 GeV. It was later summarized as the standard model of particle physics together with the QCD, based on the gauge group of SU(3)×SU(2)×U(1).

2.5 Similarities and Differences Among Four Forces

To arrive at a unified description of these forces, we need to identify the similarities and differences among them. One notable fact is that all four forces are based on gauge principles: general coordinate transformation (gravity), abelian gauge invariance (electromagnetism) or non-abelian gauge invariance (strong and electroweak forces). It is this fact which lead to the idea that all four forces may have a unified description. The ranges of forces are quite diverse, but they are understood as a consequence of different phases which each gauge theories have: Coulomb (gravity and electromagnetism), confinement (strong) and Higgs (electroweak) phases. There are, however, two major differences between gravity and the other three. (1) Gravity is related to the space-time geometry described by a spin-two field, the space-time metric. On the other hand, the other gauge fields are related to the internal space, such as phase (U(1)), weak isospin (SU(2)) or color (SU(3)) degrees of freedom, and are described by vector-fields which carry spin one. (2) The strength of gravity is described by the Newton constant with mass dimension negative two while the other three are described by dimensionless constants, such as the fine-structure constant in the natural units $\hbar = c = 1$. These two points lead to a profound difference in the quantum field theory. Theories with constants with negative mass dimensions are not renormalizable, *i.e.* there are irremovable infinities in the theory, while the ones with dimensionless coupling constants are. This fact makes the unification of electromagnetic, strong and weak forces relatively easy from field theoretical point of view while the incorporation of gravity is rather difficult.

It must be noted that there are other kinds of forces in the standard model of particle physics. There is a self-interaction of the Higgs boson which generates the mass of the Higgs boson by its own condensate, and many Yukawa interactions of the quarks and leptons to the Higgs boson which generates their masses. They are all characterized by dimensionless coupling constants. They are, however, not based on gauge principle by any means, and such interactions can be introduced with arbitrary strengths opposed to the universal character of the gauge forces. They are often not called as "forces" because we have not yet detected the effect of such interactions experimentally. However they are present in the standard model and need to be understood as well if there is unified description of *all* forces. In particular, the Yukawa interactions have a rather baroque pattern of strengths in order to reproduce the wide range of quark and lepton masses, *e.g.*, 0.511 MeV/c^2 for the electron to 175 GeV/c^2 for the top quark discovered in 1995 (CDF

Collaboration, 1995; D0 Collaboration, 1995). A truly unified theory must also explain the pattern of Yukawa interactions.

2.6 Hints for Unification

Finally, we list two puzzles in the standard model which suggest a need for a possible unified description. In the standard model, the quarks and leptons appear completely independent from each other, and different forces (namely SU(3), SU(2), and U(1) forces) are a priori not related to each other. However, then, it is not clear in the standard model why the electric charges of both quarks and leptons come in the integer multiples of 1/3. Such a quantization is essential to guarantee the electric neutrality of matter, which balances the charge of a proton made up of two up- and one down-quarks with the charge $(2/3)\times 2+(-1/3) = 1$, and the charge of an electron -1. The neutrality is tested at 10^{-21}. Second, due to quantum effects, the gauge invariance could be lost (called anomaly) which potentially spoils the unitarity or renormalizability of the theory. In the standard model, however, this does not occur due to a non-trivial cancellation. For instance, it is necessary that the sum of all electric charges vanish in the standard model particle content, which is true only after adding contribution of the quarks *and* leptons, *and* with the correct multiplicities of color (3 for SU(3)), *and* the correct multiplicities for left-handed particles (2 for SU(2)). Other necessary conditions are also satisfied thanks to contributions of both quarks and leptons with correct multiplicities among various groups. This observation suggests that different gauge groups are related at a deep level, so are quarks and leptons.

3 Grand Unified Theories

If one leaves out gravity, the unification of other three forces is not difficult because one can maintain the renormalizability of the quantum theory (for a detailed treatment, see Ross, 1984).

3.1 Original GUTs

The earliest model towards the unification of the electromagnetic, weak and strong forces is probably the one by Pati and Salam (1973), where a gauge group of SU(4)×SU(2)×SU(2) was employed. Later, Georgi and Glashow (1974) proposed a model based on a simple group SU(5), which became the reference group in almost all subsequent efforts. It was a remarkable observation that all gauge forces and quite complicated quantum numbers of quarks and leptons in the standard model fit nicely into a single gauge group SU(5) (see Table. 2). Subsequently, Georgi, Quinn and Weinberg (1974) showed that the strong force becomes weaker at higher energies while the electromagnetic force stronger, so that they have the same strength at a certain high energy: around 10^{14} GeV in the simplest example. This observation leads to the following picture of a possible unification of forces. The gauge forces (except gravity) are ultimately described by a simple group such as SU(5) (or SO(10), E_6 as described below) with a single coupling constant. It breaks spontaneously to SU(3)×SU(2)×U(1) gauge group at an extremely high energy scale (unification scale) such as 10^{14} GeV and the gauge coupling constants of different groups run differently between the unification scale and the weak scale 250 GeV. A simple calculation shows that the coupling con-

Table 2: The quantum numbers of the quarks and leptons under the standard model gauge groups and the way they can be embedded into representations under the grand unified gauge groups. In the first row, the first number is the number of color degrees of freedom under QCD, the second whether the particle couples to the electroweak group as doublets ($\mathbf{2}$) or not ($\mathbf{1}$), and the last the charge under the U(1) hypercharge group. In the latter two rows, the numbers indicate the dimensionality of SU(5) or SO(10) representations. The subscripts L, R refer to left- or right-handed helicity states.

	Q_L	u_R	e_R	d_R	L_L	ν_R
SU(3)×SU(2)×U(1)	$(\mathbf{3}, \mathbf{2}, 1/6)$	$(\mathbf{3}, \mathbf{1}, 2/3)$	$(\mathbf{1}, \mathbf{1}, -1)$	$(\mathbf{3}, \mathbf{1}, -1/3)$	$(\mathbf{1}, \mathbf{2}, -1/2)$	$(\mathbf{1}, \mathbf{1}, 0)$
SU(5)	**10**			$\mathbf{5}^*$		**1**
SO(10)	**16**					

stant becomes larger at lower energies for a larger gauge group. If one follows the running from the unification scale down to the weak scale, the coupling constant of a larger gauge group become larger than that of a smaller gauge group. Therefore, this picture, called grand unified theory or GUT, explains why the strong force is strong and the electromagnetic force is weak simply based on the size of the groups, $3 > 2 > 1$. Furthermore, quantization of the electric charge is explained as a simple consequence of the group theory.

An interesting feature of SU(5) unification is that the quarks and leptons belong to the same multiplets under SU(5), so that they can transform into each other by the exchange of SU(5) gauge bosons, in the same way that an up-quark and down-quark can transform into each other through the charged-current weak interaction mediated by the exchange of the W-boson.

Therefore, the SU(5) unification predicts that a quark can turn into a lepton, which allows a proton to decay (*e.g.* $p \rightarrow e^+\pi^0$ in the simplest model), a new phenomenon that can be looked for by experiments. Such a violation of baryon number may explain the reason why the Universe contains matter (protons, neutrons and electrons) but no anti-matter (anti-protons, anti-neutrons and positrons) as a result of baryon-number-violating interactions which occurred when the Universe was as hot as the unification energy scale (baryogenesis) (Yoshimura, 1978; Ignatiev *et al*, 1978). Finally, it was also pointed out that the masses of the bottom quark and tau lepton, which differ roughly by a factor of three in laboratories, also become the same size when extrapolated up to a similar high energy scale. This observation led to a hope that the pattern of quark and lepton masses could be also understood in the context of unified theories.

Unfortunately, the original SU(5) unified model is now excluded because of two reasons. First, the non-observation of proton decay, especially by the IMB and Kamiokande experiments, has put lower limits on the proton life time which exceed the prediction of the model. Second, the LEP and SLC experiments measured the size of three gauge coupling constants for three gauge groups SU(3)×SU(2)×U(1) accurately, which did not agree with the predicted pattern in the model. There were other problems pointed out concerning the original SU(5) unified model. The assumption that there are two very disparate energy scales (hierarchy), the unification scale and the weak scale, appears unnatural, which will be discussed more below. The cancellation of the anomaly is still not obvious because one needs to add contributions of two multiplets ($\mathbf{5}^*$ and $\mathbf{10}$) under SU(5). The theory predicts the existence of magnetic monopoles of a mass higher than the unification scale, which have never been detected. The unification of quark and lepton masses works for the third generation (bottom and tau), but not so well for first and second generations.

The question of anomaly cancellation led people to consider yet larger groups whose irreducible multiplet contains the whole generation which belonged to two irreducible multiplets under SU(5). One such candidate is SO(10) which carries certain unique features (see Table 2). It is the smallest group which allows chiral fermions while automatically anomaly-free. It predicts another particle, the right-handed neutrino, for each generations. By allowing the right-handed neutrinos to acquire masses at a high scale, such as unification scale, they become irrelevant to typical experiments. However, the diagonalization of neutrino mass matrix leads to a small but finite masses for the left-handed neutrinos, given approximately by $m_\nu \sim m_u^2/M$ where M is the mass of the right-handed neutrinos and m_u is the mass of up-type quark in the same generation. Therefore, a large M results in a small m_ν well below eV$/c^2$ (Yanagida, 1979; Gell-Mann *et al.*, 1979), which may explain the observed deficit in the solar neutrino flux, a part of the missing mass in the Universe (Hot Dark Matter), or various other observations in neutrino experiments (see Fukugita and Yanagida, 1994, for a review on neutrino physics). People discussed other anomaly-free gauge groups along the same line, such as E_6, or $(\mathrm{SU}(3))^3$.

3.2 Supersymmetric GUTs

The existence of a huge hierarchy between the unification scale and the weak scale posed a serious theoretical challenge. In particular, it was found that the correction to the Higgs boson mass squared from higher orders in perturbation theory is quadratically divergent, and tends to either bring the weak scale up to the unification scale or restores the electroweak group SU(2)×U(1) (no Higgs mechanism). A mechanism to stabilize the hierarchy in energy scales seemed necessary (see Murayama, 1994, for a review on this discussion). Supersymmetry (see Wess and Bagger, 1983, for a textbook) was postulated as such a mechanism. It is well-known that the electron acquires a linearly divergent Coulomb self-energy in classical electrodynamics. If one extrapolates the electrodynamics to a much shorter distance scale, the Coulomb self-energy grows linearly with the distance scale and so is the electron mass. However the existence of positron and its appearance as a part of quantum fluctuation of vacuum cancels the linearly divergent Coulomb self-energy and the result is only logarithmically divergent for small distances. A symmetry, called chiral symmetry, which interchanges an electron with a positron, is responsible for the cancellation of wild behavior of the self-energy. After such a cancellation, the quantum electrodynamics can be applied to much shorter distance scales, or equivalently, much higher energy scales without ruining the smallness of the electron mass. In a similar fashion, supersymmetry introduces *superpartners* to each particles in the standard model, whose spins differ by 1/2 and have opposite statistics (Bose vs Fermi). And their contribution in the quantum fluctuation cancels the quadratically divergent contribution to the Higgs boson mass squared and makes the standard model applicable up to the unification scale without spoiling the smallness of the weak scale.

Therefore, grand unified theories are often discussed with a supersymmetric extension (Dimopoulos and Georgi, 1981; Sakai, 1981). The precise determination of the three gauge coupling constants at LEP and SLC experiments running from 1989 revealed that they seem to unify remarkably well at an energy scale of 2×10^{16} GeV if we assume the minimal supersymmetric extension of the standard model, while they do not unify at all with a non-supersymmetric standard model (see Fig. 1). This observation led to a renewed enthusiasm for supersymmetric grand unified theories. Superpartners of the standard model particles are currently under intensive searches at various collider experiments, such as the upgraded LEP and Tevatron, and are expected to be found at latest at the LHC experiments which begin data taking in 2005 (Baer *et al.*, 1995). The proton lifetime is expected to be within the reach of superKamiokande experiment which started in 1996, in the simple models (Murayama, 1996). Other experimental signatures of supersymmetric grand unified theories were pointed out, such as $\mu \to e\gamma$ (Barbieri and Hall, 1994). They are under active experimental efforts as well. Once the superparticles are found, there are numerous other tests possible on grand unified theories. The mass spectrum of superparticles has a definite pattern if the forces are grand unified. By measuring the masses of superparticles, one can either verify or exclude grand unified theories, which is possible at a percent level at future collider experiments (Tsukamoto *et al.*, 1995; Hinchliffe *et al.*, 1996).

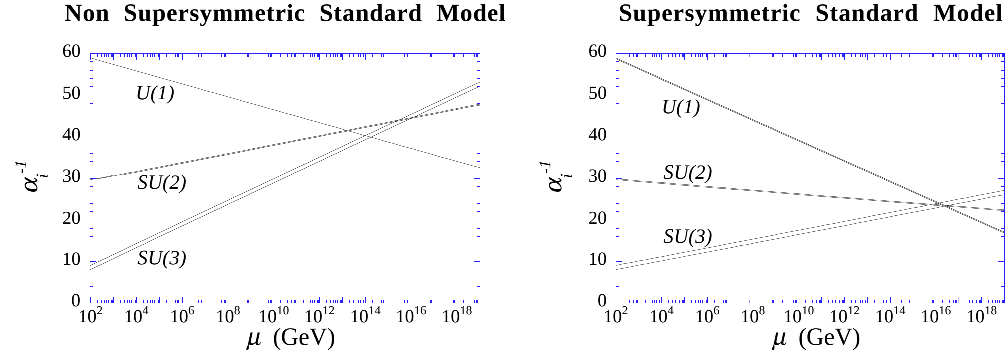

Figure 1: The running of three gauge coupling constants in the standard model, assuming the minimal non-supersymmetric standard model or its supersymmetric extension.

The supersymmetric grand unified theories are not without problems, however. Most importantly, supersymmetry *stabilizes* the hierarchy between the unification scale and the weak scale, but does not explain why such a hierarchy exists to begin with. One common scheme assumed in literature is that supersymmetry is broken due to a strong dynamics of a gauge theory, analogously to the chiral symmetry breaking due to the strong QCD dynamics, which is the origin of the masses of proton and neutron. Such a scheme naturally generates spontaneous supersymmetry breaking at an energy scale $\simeq M_{\text{Planck}} e^{2\pi/\alpha b_0}$ much lower than the unification scale or Planck scale, if the gauge coupling constant α is perturbative at the Planck scale, and if the gauge theory is asymptotically free $b_0 < 0$. Furthermore, the effect of broken supersymmetry is mediated to the standard model particles only by gravitational interactions (Chamseddine *et al.*, 1982; Hall *et al.*, 1983). Therefore, incorporation of gravity into a unified theory appeared necessary.

Supersymmetry actually suggests a possibility to relate the gauge forces to gravity whose force carriers have spin 1 and 2, respectively. Since supersymmetry relates particles with different spins, they may be related as well. Furthermore, supersymmetry can be promoted to a gauge invariance, *local* supersymmetry, which necessarily incorporates gravity (supergravity). Finally, supersymmetry makes the theory behave more reasonably at high energies, so that gravity, which has a severe problem of divergences at high energies and hence is non-renormalizable, may be better behaved at the quantum level once it is made supersymmetric.

Supergravity itself does not make the quantum theory of gravity fully consistent. It still is non-renormalizable. In early 1980's there was a speculation that supergravity with the maximal supersymmetry (de Wit and Freedman, 1977; Cremmer and Julia, 1978; 1979; de Wit, 1979) (called $N = 8$ because it possesses 8 times more supersymmetries than the smallest supergravity theory) may be finite and hence consistent. It also had a large internal symmetry which allowed 28 gauge bosons (de Wit and Nicolai, 1981; 1982; Hull, 1984), and, people hoped, may contain the standard model gauge groups. Unfortunately, neither of the expectations turned out to be true and this direction of research has been basically abandoned. Possible counter terms were found at the seven-loop level (Howe and Lindstrom, 1981) which indicated that the theory has ultraviolet divergences. It was difficult to embed the standard model gauge groups into the internal symmetry of the theory. And people could not embed chiral matter content into the theory. Supergravity is now regarded as an effective description of a true quantum theory of gravity valid only up to the energy slightly below the Planck scale.

Recently, a mechanism to mediate the effect of supersymmetry breaking to the standard model particles by usual gauge forces has attracted interests (Dine *et al.*, 1993, 1995, 1996). In such a case, supergravity does not appear crucial for the purpose of generating supersymmetry breaking effects which are phenomenologically necessary. The mechanism has a phenomenologically appealing feature that the quantum contribution of superparticles to rare phenomena are naturally suppressed below the experimental limits. There is ongoing active research to understand the phenomenological consequences of this mechanism (Dimopoulos *et al.*, 1996; Babu *et al.*, 1996). It appears, however, relatively contrived at the moment, and a substantial effort is being devoted to simply it. Discussion of phenomenology of grand unified models with this mechanism has just started (Carone and Murayama, 1996).

4 Kaluza–Klein Theories

Soon after the Einstein's general theory of relativity, people tried to formulate electromagnetism on an equal footing. One possible direction is to obtain the electromagnetic gauge invariance (U(1)) as a part of the general coordinate transformation, which requires the extension of our four-dimensional space-time to higher dimensions. Such theories are called Kaluza–Klein theories according to the authors of early works (Kaluza, 1921; Klein, 1926). In general, this class of theories start with a higher dimensional space-time with a certain theory of gravity, and try to obtain other gauge forces in the standard model upon "compactifying" extra dimensions.

The simplest example by Kaluza and Klein starts with a five-dimensional space-time. One "compactifies" the extra spatial direction to a small circle, and the rest large four dimensions are identified with our ordinary space-time. The metric tensor of the five-dimensional space-time $G_{\mu\nu}$, $\mu, \nu = 0, 1, 2, 3, 4$ now decomposes into the four-dimensional metric $g_{\mu\nu}$ $(\mu, \nu = 0, 1, 2, 3)$, a vector $A_\mu = g_{4\mu}$, and a scalar $S = g_{44}$. The original five-dimensional general coordinate transformation acts as the ordinary gauge transformation on the vector A_μ. In general, an isometry of the compactified space generates an effective gauge invariance in the remaining four-dimensional space-time. Much later, it was found that such a space-time could be a solution to the (modified) Einstein's equation that it could be regarded as a "spontaneous" compactification of a high-dimensional space-time.

This simple and beautiful idea suffers from all difficulties which the Einstein's general relativity has and more. First of all, quantum theory is not well-defined because quantum theory of gravity is yet-to-be understood. Second, electromagnetism, which can be quantized and studied as a renormalizable quantum field theory, becomes non-renormalizable in higher dimensions. Third, the five-dimensional theory cannot generate gauge fields for strong and weak forces (see for a review and a collection of papers, Appelquist *et al.*, 1985).

Witten proposed a "realistic" version of the Kaluza–Klein theory (Witten, 1981). In order to incorporate all known forces, strong, weak, and electromagnetic, the space-time dimension has to be equal to or larger than 11. On the other hand, supersymmetry, as explained in the previous section, is probably an essential ingredient in unified theories. A space-time higher than 11 dimensions generates too many supersymmetry generators in four dimensions after compactifications, which result in theories with unwanted higher spin fields, such as spin 5/2 or 3, whose existence would make the theory behaved even worse at high energies. Therefore, 11 is the only dimensionality which may be realistic. In fact, solutions to Einstein's equation were found which allow compactified seven-dimensional space with SU(3)×SU(2)×U(1) isometry (Castellani *et al.*, 1984). This observation aroused a great enthusiasm in 11-dimensional supergravity theories (Cremmer *et al.*, 1978). However, the uniqueness of the dimensionality posed a serious difficulty in pursuing the direction further. Witten later showed (Witten, 1983) that such a theory can never produce chiral fermions in four dimensions. Therefore, at least some extra ingredients (*i.e.*, gauge fields) are needed before compactifications to fully incorporate all known forces on equal footing. The apparent absence of the cosmological constant in our four-dimensional space-time was also very difficult to obtain in this framework. Finally, it did not solve the problem concerning the non-renormalizability of the theory.

The Kaluza–Klein theories, at least with the original framework, do not seem to offer a successful unified description of all four forces. It is noteworthy, however, that eleven-dimensional supergravity theories have renewed interests in the context of string duality which will be discussed later.

5 Superstring Theory

The (partial) success of supersymmetric grand unified theories led to an optimism on the basic concept of unification. There is no compelling and aesthetically appealing concrete model at hand, but the concept itself appears promising. On the other hand, the incorporation of gravity has been a serious challenge even with supersymmetry because of bad behavior of the quantum gravitational effects at high energies.

The string theory (see Green *et al.*, 1987 for a textbook) replaces particles by small strings with a typical size of the Planck length, $\sim 10^{-33}$ cm. The ordinary particles, with masses much less than the Planck mass, are identified with the lowest excitation modes of the string. Historically string model of particles was developed by Nambu and Goto to explain properties of hadrons which were not understood by then (see Scherk, 1975, for a review). Later, however, it became clear that the string theory contains massless spin-one and spin-two excitations which can potentially be identified with gauge bosons and graviton. In particular, the equations of motion for spin-two field derived from the string theory agreed with the Einstein's theory of general relativity at low-energy approximation. Therefore, the focus of string theory changed to a consistent framework of quantum gravity (Scherk and Schwarz, 1974).

Green and Schwarz (1984) found that certain classes of string theory are free from anomalies and are ultraviolet finite. This observation aroused the hope that the string theory provides a consistent framework of quantum gravity. They were classified into the following categories: type-I, type-IIA, type-IIB, and heterotic. The heterotic string theory (Gross *et al.*, 1985) allowed only two choices for the gauge group, SO(32) or $\mathrm{E}_8\times\mathrm{E}_8$, and the latter can incorporate the standard model or grand unified groups such as SU(5), SO(10) or E_6. The other E_8 can act as the sector which breaks supersymmetry well below the Planck scale. The theories could be formulated only in ten-dimensional space time. It was further shown, however, that extra six dimensions can be compactified as a solution to the equations of motions, on so-called Calabi–Yau manifolds or orbifolds, such that the remaining four-dimensional space time has minimal supersymmetry and realistic gauge groups and chiral matter content (Candelas *et al.*, 1985). Furthermore, a string theory contains only a single coupling constant. If the theory is supposed to describe particle physics, it is supposed to explain *all* coupling constants by a single parameter. Therefore, if string theory is correct, it must be the "Theory of Everything."

There are a number of problems in string theory, however, at the moment. One of the major problems is that the string theory allows too many solutions to its equations of motion, which give different matter content, different gauge groups, different coupling constants, and even different space-time dimensions. It was hoped that non-perturbative effects would eliminate most of the solutions and could pick the unique solution which describes our Universe. It requires a non-perturbative formulation of string theory which is still not available. String theory is formulated only in the context of perturbation series at this moment. Phenomenological issues, such as whether the string theory is consistent with the observed pattern of gauge coupling constants (for a recent compilation of discussions, see Dienes *et al.*, 1996), are discussed mostly within perturbation theory.

The invention of string duality (Hull and Townsend, 1995; Witten, 1995) is changing the situation drastically. The behavior of string theory in non-perturbative regime is now understood much better by using the duality transformation, which maps a strongly-coupled regime of one string theory to a weakly-coupled regime of another string theory. This is a generalization of electric-magnetic duality in electromagnetism or duality in two-dimensional Ising model which interchange small and large coupling constants. Much more non-trivial situation was revealed in supersymmetric gauge theories (for a review, see Intriligator and Seiberg, 1996). Similar techniques were employed in string theory which led to discoveries of string duality. In particular, the heterotic string theory with $\mathrm{E}_8 \times \mathrm{E}_8$ gauge group, which appears to be the most promising one phenomenologically, is now believed to be dual to the 11-dimensional supergravity with a rather peculiar topology of the eleventh direction (called M-theory) (Hořava and Witten, 1996). The theory contains no parameters at the Lagrangian level. The eleventh dimension is compactified on a finite interval $x^{11} \in [0, \rho]$, and the interval ρ plays the role of the coupling constant of the theory. All the gauge and matter fields live only inside the ten-dimensional hyperplanes at the boundaries $x^{11} = 0$ and $x^{11} = \rho$, while the supergravity multiplet can propagate both in ten-dimensional and eleventh direction. Since there are gauge fields already present before the compactification to four dimensions, the theory is free from the problem of chiral fermions discussed before in the context of Kaluza–Klein theories. There are indications that the M-theory description is the most appropriate one for the quantum gravity and gauge forces among the elementary particles. Yet many problems are left unsolved, especially on the uniqueness of the solution.

6 Summary

There has been many attempts to have a unified description of all observed forces in Nature. Different attempts have different levels of success. Among them, supersymmetric grand unified theories appear to be the most promising phenomenological models which offer us a unified description of electromagnetic, strong, and weak forces. Much experimental efforts are being devoted to either detect superparticles or rare phenomena predicted by grand unified theories, and some of them are within the reach of the near future experiments. The incorporation of gravity, however, requires a formulation of quantum gravity which is impossible within the standard quantum field theory. The most promising candidate formulation of quantum gravity is the string theory, and its no-perturbative dynamics has just begun to be understood. Concrete phenomenological predictions are yet to be worked out.

Acknowledgments

HM thanks Christopher Carone for careful reading of and comments on the manuscript. This work was supported in part by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics of the U.S. Department of Energy under Contract DE-AC03-76SF00098, in part by the National Science Foundation under grant PHY-90-21139, and also by Alfred P. Sloan Foundation.

List of Works Cited

Appelquist, T. *et al.*(1985), (Ed.), Modern Kaluza–Klein theories, Menlo Park: Benjamin/Cummings.

Babu, K.S. *et al.*(1996), Phys. Rev. Lett. 77, 3070.

H. Baer, *et al.*(1997), in: Barklow, T. L. *et al.*, (Ed.), Electroweak Symmetry Breaking and New Physics at the TeV Scale, Singapore: World Scientific, .

Barbieri, R., Hall, L.J. (1994), Phys. Lett. B338, 212.

Bloom, E.D., *et al.*(1969), Phys. Rev. Lett. 23, 930.

Breidenbach, M. *et al.*(1969), Phys. Rev. Lett. 23, 935.

Candelas, P. *et al.*(1985), Nucl. Phys. B258, 46.

Carone, C. D., Murayama, H. (1996), Phys. Rev. D53, 1658.

Castellani, L. *et al.*(1984), Nucl. Phys. B239, 610.

CDF Collaboration (1995), Phys. Rev. Lett. 74, 2626.

Chamseddine, A.H. *et al.*(1982), Phys. Rev. Lett. 49, 970.

Cremmer, E. B. *et al.*(1978), Phys. Lett. 76B, 409.

Cremmer, E., Julia, B. (1978), Phys. Lett. 80B, 48.

Cremmer, E., Julia, B. (1979), Nucl. Phys. B159, 141.

Dienes, K. R. *et al.*(1996), Nucl. Phys. B467, 44.

Dimopoulos, S., Georgi, H. (1981), Nucl. Phys. B193, 150.

Dimopoulos, S. *et al.*(1996), Phys. Rev. Lett. 76, 3494.

Dine, M., Nelson, A.E. (1993), Phys. Rev. D48, 1277.

Dine, M. *et al.*(1995), Phys. Rev. D51, 1362.

Dine, M. *et al.*(1996), Phys. Rev. D53, 2658.

D0 Collaboration (1995), Phys. Rev. Lett. 74, 2632.

Fukugita, M., Yanagida, T. (1994), in: Fukugita M., Suzuki, A. (Ed.), Physics and Astrophysics of Neutrinos, Tokyo: Springer-Verlag, p. 1.

Gell-Mann, M.*et al.*(1979), in: van Nieuwenhuizen, P., Friedman, D.Z. (Ed.), Supergravity, Amsterdam: North Holland.

Georgi, H., Glashow, S.L. (1974), Phys. Rev. Lett. 32, 438.

Georgi, H., Quinn, H.R., Weinberg, S. (1974), Phys. Rev. Lett. 33, 451.

Glashow, S.L. (1961), Nucl. Phys. 22, 579.

Green, M. B., Schwarz, J. H. (1984), Phys. Lett. 149B, 117.

Green, M. B., Schwarz, J. H. (1985), Phys. Lett. 151B, 21; Nucl. Phys. B255, 93.

Green, M. B. *et al.*(1987), Superstring Theory, Cambridge: Cambridge University Press.

Gross D.J., Wilczek, F. (1973), Phys. Rev. Lett. 30, 1343.

Gross, D. J. *et al.*(1985), Phys. Rev. Lett. 54, 502; Nucl. Phys. B256, 253.

Hall, L. *et al.*(1983), Phys. Rev. D27, 2359.

Hasert F.J. *et al.*(1973), Phys. Lett. B46, 138.

Hinchliffe, I. *et al.*(1997), Phys. Rev. D55, 5520-5540.

Horava, P., Witten, E. (1996), Nucl. Phys. B460, 506; B475, 94.

Howe, P., Lindstrom, U. (1981), Nucl. Phys. B181, 487.

Hull, C. M. (1984), Phys. Rev. D30, 760; Phys. Lett. 142B, 39.

Hull, C.M., Townsend, P.K. (1995), Nucl. Phys. B438, 109.

Ignatiev, A. Yu. *et al.*(1978), Phys. Lett. 76B, 436.

Intriligator, K., Seiberg, N. (1996), Nucl. Phys. Proc. Suppl. 45BC, 1.

Kaluza, Th. (1921), Sitzungsber, Preuss. Akad. Wiss., Berlin, Math. Phys. K1, 966 .

Kinoshita, T. (1995), in: Newman, H., Ypsilantisin, T. (Ed.), History of Original Ideas and Basic Discoveries in Particle Physics, New York: Plenum Press, .

Klein, O. (1926), Z. Phys. 37, 895.

Murayama, H. (1995), in: Yamada, S., Ishii, T. (Ed.), 22nd INS International Symposium on Physics with High Energy Colliders, Singapore: World Scientific.

Murayama, H. (1997), in: Ajduk, Z., Wroblewski, A. K. (Ed.), 28th International Conference on High-energy Physics (ICHEP 96), Singapore: World Scientific.

Particle Data Group (1996), Phys. Rev. D54, 1.

Pati, J.C., Salam, A. (1973), Phys. Rev. D8, 1240.

Politzer, H.D. (1973), Phys. Rev. Lett. 30 1346.

Ross, G. G. (1984), Grand Unified Theories, Menlo Park: Benjamin/Cummings,.

Sakai, N. (1981), Zeit. Phys. C11, 153.

Salam, A. (1968), in: Svartholm (Ed.), Elementary Particle Theory, Proceedings Of The Nobel Symposium, Sweden, Stockholm.

Scherk, J. (1975), Rev. Mod. Phys. 47, 123.

Scherk, J., Schwarz, J. H. (1974), Nucl. Phys. B81, 118.

Tsukamoto, T. *et al.*(1995), Phys. Rev. D51, 3153-3171.

Weinberg, S. (1967), Phys. Rev. Lett. 19, 1264.

Wess, J. Bagger, J. (1983), Supersymmetry and Supergravity, Princeton: Princeton University Press.

de Wit, B., Freedman, D. Z. (1977), Nucl. Phys. B130, 105.

de Wit, B. (1979), Nucl. Phys. B158, 189.

de Wit, B., Nicolai, H. (1981), Phys. Lett. 108B, 285

de Wit, B., Nicolai, H. (1982), Nucl. Phys. B208, 323.

Witten, E. (1981), Nucl. Phys. B186, 412.

Witten, E. (1983), in: Jackiw, R. *et al.*, Shelter Island Conference on Quantum Field Theory and the Fundamental Problems of Physics, 2nd, Cambridge: MIT Press.

Witten, E., IASSNS-HEP-95-63, hep-th/9507121, contributed to STRINGS 95: Future Perspectives in String Theory, Los Angeles, CA, 13-18 Mar 1995.

Yanagida, T. (1979), in: Sawada, O., Sugamoto, A., Proc. of the Workshop on the Unified Theory and Baryon Number in the Universe, KEK report 79-18.

Yoshimura, M. (1978), Phys. Rev. Lett. 41, 281; Erratum ibid. (1979) 42, 746.

Further Reading

Appelquist, T. *et al.*(1985), (Ed.), Modern Kaluza–Klein theories, Menlo Park: Benjamin/Cummings.

Close, F. (1983), The Cosmic Onion: quarks and the nature of the universe, Heinemann.

Green, M. B. *et al.*(1987), Superstring Theory, Cambridge: Cambridge University Press.

Georgi, H. (1982), Lie Algebras in Particle Physics: from isospin to unified theories, Menlo Park: Benjamin/Cummings,.

Mohapatra, R. N. (1992), Unification and Supersymmetry: the frontiers of quark-lepton physics, Springer-Verlag.

Nilles, H.P., (1984), Phys. Rept. 110, 1,.

Ross, G. G. (1984), Grand Unified Theories, Menlo Park: Benjamin/Cummings,.