The future of unified gauge theories

Having unified the weak and electromagnetic and perhaps the strong interactions, gauge theories now point towards key roles for Higgs bosons, and possible particles with masses as high as 10^{17} GeV.

Steven Weinberg

Unified gauge theories of elementary-particle interactions have had an exciting start. In this article, I want to try to look a little way into their future.

To begin with, it may be helpful to remind the reader of a few of the chief points of these theories. The subject has been reviewed very often, so I will be brief; for additional information the reader is referred to the reading list at the end of the article. The Box on page 47 contains a glossary of some of the more technical terms used here.

▶ Gauge theories are characterized by their invariance under a group of symmetry transformations. The transformations are "internal," like isotopic-spin "rotations," acting on the labels of particles rather than on their space–time coordinates. However, unlike isospin rotations, the effect of these transformations is supposed to depend on position in space–time. The archetypal gauge theory is electrodynamics, for which the gauge group is U(1), the group of space–time-dependent phase transformations on charged fields.

▶ Gauge invariance requires that the theory involves a set of massless vector fields, such as the photon field. In unified gauge theories, these fields are supposed to mediate not only the electromagnetic interactions but also the weak interactions responsible for nuclear beta decay, and perhaps as well the strong interactions, which hold atomic nuclei together.

Steven Weinberg is a visting professor at the Stanford University Physics Department while on leave from his positions as Higgins Professor of Physics at Harvard University and Senior Scientist at the Smithsonian Astrophysical Observatory. On 8 February he was awarded the Dannie Heineman Prize for Mathematical Physics; this article is an adaptation of his talk on that occasion.

▶ The gauge symmetries are supposed to be truly exact symmetries of the underlying field equations and commutation relations. However, part of this gauge group is spontaneously broken; that is, it is not realized in physical states. This gives a mass to the vector bosons associated with the broken symmetries and accounts for the obvious observed differences between the weak and electromagnetic interactions. This spontaneous symmetry breakdown is similar to the loss of translational symmetry when liquid water turns into an ice crystal lattice below 0°C. Indeed, the apparent difference between weak and electromagnetic interactions can be similarly blamed on the circumstance that the Universe is now quite cold.

▶ The gauge theories of particular physical interest are of the simple, highly constrained type called "renormalizable." In such theories, all infinities appearing in perturbation theory can be absorbed into a redefinition of physical parameters such as masses and coupling constants.

My remarks will be divided into three parts, corresponding to three ranges of energy, which I will call simply low, medium and high (but not with the meaning that ERDA gives to these terms). By *low* energy I mean here the range accessible to us until now and over the next few years, involving particles with masses from zero up to a few GeV. By *medium* energy I mean the range we will be exploring for the next few decades, involving masses up to a few hundred GeV. One of the new accelerators that will open up the medium energy range is the Positron–Electron Tandem Ring Accelerator shown under construction in figure 1. *High* energy is what lies beyond, perhaps extending up to 10^{19} GeV.

The direct experimental study of the high energy range may not be within the economic resources of the human race, at least until someone invents something very clever. However, in our present studies of the low energy range, we are already forced to speculate on what happens at high energy. As I will try to show, the theory of the medium energy range plays an even more essential part in our understanding of what we now observe.

Low energy

It now appears that the weak, electromagnetic and strong interactions that we observe at present accelerator energies are governed by a renormalizable gauge theory, analogous to quantum electrodynamics. Instead of the one-dimensional gauge invariance group, U(1), of pure electrodynamics, the gauge group of the weak and electromagnetic interactions is the four-dimensional group SU(2) ⊗ U(1), or perhaps something a little larger. (I use here a standard notation: SU(N) is the group of unitary $N \times N$ matrices with unit determinant.) This gauge symmetry is supposed to be strongly spontaneously broken, accounting for the obvious differences between the weak and electromagnetic interactions. The gauge group of the strong interactions is very likely the "color" group SU(3) and is probably not broken at all. In addition to the gauge vector bosons (the photon, $W^{\pm}$, Z^0 for SU(2) ⊗ U(1), and, for SU(3), eight colored "gluons") the world contains leptons, which are SU(3) neutral, and colored quarks, which form SU(3) triplets and anti-triplets. Figure 2 lists these fields and gives some of their properties.

I do not mean to imply that everyone agrees with every part of this picture. There are distinguished theorists who disagree with some of its aspects, particularly with regard to the strong interactions. However, it is useful to have a detailed picture available, if only to have something definite to test. And in its general outlines, this picture agrees with

Tunnel construction is nearly completed at the PETRA electron–positron ring in Hamburg (see "Search and Discovery," PHYSICS TODAY, June 1976, page 17 for details). What new particle data will PETRA, PEP (being built at the Stanford Linear Accelerator Center) and other new accelerators produce? How will these data fit in with the gauge theories? (Photo courtesy of G.-A. Voss, DESY) Figure 1

so many of our theoretical requirements and with so much experimental information that it must be essentially correct.

The most dramatic confirmations of the general gauge picture of weak, electromagnetic and strong interactions have of course come from the discovery of neutral currents at CERN and at Fermilab in mid-1973, and also the discoveries of new hadron states, starting with "J/ψ" at Brookhaven and SLAC in late 1974.

However, it is important to keep in mind not only the *predictions* of the gauge theories, but also their *retrodictions:* the many things that have been known empirically for years and that the renormalizable gauge theories have now explained. I will list some of these, taking first the retrodictions having to do with the strong interactions alone, and then those that involve the weak and electromagnetic interactions as well:

Partons and scaling As long as we do not allow strongly coupled scalar fields or too many quarks, the SU(3) gauge theory of strong interactions has the property of "asymptotic freedom," that the strong interactions become weaker at high energy. This accounts for at least some of the features of the parton model, particularly the Bjorken scaling observed in deep inelastic electron–nucleon reactions.

Quark trapping(?) The other side of asymptotic freedom is infrared slavery, the idea that the strong interactions become so strong when quarks or gluons are widely separated that neither quarks nor gluons can be isolated as free particles. A mathematical demonstration of infrared slavery is still lacking, and of course quarks may yet be found, so I list this item with a question mark.

Exact symmetries of strong and electromagnetic interactions Renormalizable gauge theories of the strong interactions that do not involve strongly coupled scalar fields are powerfully constrained by the conditions of gauge invariance and renormalizability—so much so that the only free parameters in such theories are a single gauge coupling constant and the elements of the mass matrix of the quarks. In consequence, these theories automatically satisfy a number of conservation laws (at least in perturbation theory), even if these laws are not assumed from the outset. These include parity conservation (P), charge-conjugation invariance (C), time-reversal invariance (T) and conservation of quark "flavors," such as strangeness (S). The same is true of the electromagnetic interactions.

Approximate symmetries of strong interactions In addition to the above "exact" symmetries, the strong interactions can have a number of approximate symmetries, depending on the properties of the quark mass matrix. If this matrix has n eigenvalues that are approximately equal, then the strong interactions have an approximate SU(n) invariance; if it has n eigenvalues that are small (whether or not their ratios are near unity), then the strong interactions have an approximate SU(n) ⊗ SU(n) invariance.

Experimentally, we know that the strong interactions have a fairly good SU(2) symmetry (isospin conservation) and a more approximate SU(3) symmetry (the "eight-fold way"), while the success of the soft-pion theorems indicates that these are subgroups of larger spontaneously broken symmetry groups, chiral SU(2) ⊗ SU(2) and SU(3) ⊗ SU(3). This pattern of symmetries is just what we would expect if there are three relatively light quarks, u, d and s, with u and d much lighter than s. (On the basis of current-algebra calculations one would estimate the u, d and s quark masses to stand in the ratio 1 : 1.8 : 36. Isospin is a good symmetry in nuclear physics not so much be-

Fields	Spin	Flavor?	Color?
Quarks	½	yes	yes
Leptons	½	yes	no
Gluons	1	no	yes
γ, $W^{\pm}$, Z^0	1	yes	no
"Higgs" particles	0	yes	no

The fundamental fields of physics that appear in gauge theories of the weak, electromagnetic and strong interactions. "Flavor" quantum numbers distinguish quarks and leptons of different charge, strangeness muon number, etc.; "color," among quarks of the same flavor. Gluons meditate the strong interactions, coupling to color; the photon and the $W^{\pm}$ and the Z^0 particles meditate the electromagnetic and weak interactions, coupling to flavor. Figure 2


AL=79.701 AR=99.187 PK=1172.9
A= 72377 CO-60 PC= 28376
Canberra.
When it counts.

Canberra MCAs. Practical. Unique.

You know Canberra as the inventive maker of multichannel analyzers—the first to replace the tedium of toggle switches with human-engineered analysis pushbuttons. You know Canberra as a pioneer of CRT Alphanumeric Displays. You know Canberra as the first to include microprocessors for standalone isotope identification. This sort of innovation and technology brings about the widest range of standalone MCAs available today. Each provides the ultimate in human engineering and state-of-the-art performance. Now see for yourself how the concept works. The OMEGA 1, an all solid state, portable multichannel analysis system has everything needed for multichannel spectrum analysis in one compact package—even an optional detector HVPS.

Small in size, OMEGA 1 is big in performance. Its CRT provides live and dynamic data display, with optional dual, digitally-controlled cursors for region of interest selection. Unique Visual Spectrum Stripping capability and Analog Compare are standard features. Optional Character Generator and Digital Integrator provide optimum data analysis. Now move up to the Canberra 8100 series. See how its features fit your own research or analysis requirements. It has the data collection capability you've been looking for—in the pulse height analysis mode, in the multichannel scaling mode, in the multiple input multichannel scaling mode, which allows up to four independent input signals. Memory? Random access 24-bit parallel BCD memory; 1024 channels semiconductor standard, 4096 semiconductor or core optional. The Input/Output Processor records data, time, number of sweeps, channel identification, and region of interest definition. Whether you select the standard instrument or any of several broad-function options, you are buying a proven system, fully developed, with a well-deserved universal popularity. Every one is cost effective, every one works.

Other features: The 5-inch rectangular CRT offers multiple regions of interest, and expansion of regions of interest, as well as roll out of expanded regions. Six display parameters can be shown simultaneously. Displays can be live or dynamic; cursor speed is digitally-controlled. Push-button digital energy calibration is optional. Pushbuttons also do indexing to the next region of interest; you can integrate full spectrum, region of interest, or expanded region. Log display is standard.

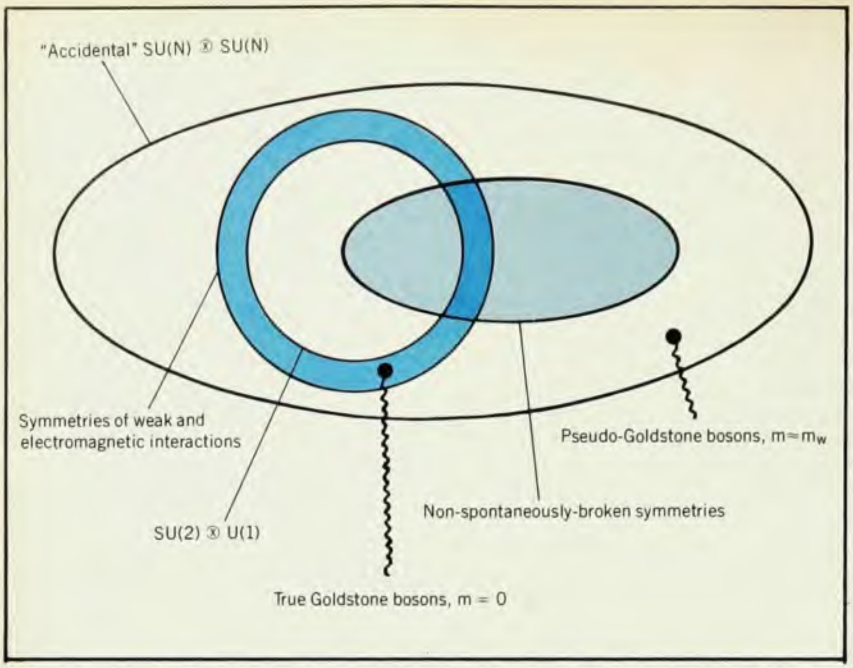
And you're in control: all standard acquisition modes—preset count, preset true or live time. Add, subtract, or non-alter data collection. You also have remote selection of all modes. Output? Data fields change automatically with display mode. You pick from printers, plotters, teletypes, magnetic tapes, NIM daisy chain modules, and EIA compatible devices. Input? From teletype, magnetic tape, cassette tape, and EIA devices. An on-line computer interface gives you real time control and analysis. The front end is just as powerful and versatile as every other element in the system. With built in amplifier and SCA, built in ADC, with a resolution of 8192 channels and excellent linearity. The memory is powerful and flexible—with 1024 or optional 4096 channels—with field expandable core or semiconductor memory modules.

A host of I/O and data analysis options increase the flexibility of 8100 series. The Canberra 8180 further expands the concept of what has been described as the smartest standalone MCA. A microprocessor for real time isotopic analysis. Computer and software power at the price and simplicity of a hardwired MCA. The read-only memory of 8K computes and displays answers, and identifies the isotope that caused the peak. An internal amplifier is a true active filter spectroscopy amp. A Pulse Pileup Rejector/Live Time Corrector can handle the highest count rates you're likely to come across. Even the I/O Processor is intelligent. It remembers the buttons you pushed. It allows you to recall the sequence at will. It's powerful stuff. Cutting right through tedious tasks, such as energy calibration, peak location, peak integration and spectrum comparison. Write for our new 20-page brochure or call us at (203) 238-2351.

CANBERRA

CANBERRA INDUSTRIES INC., 45 Gracey Ave., Meriden, CT 06450
CANBERRA INSTRUMENTS LTD., Reading, Berkshire, United Kingdom
CANBERRA INSTRUMENTATION, S.A.R.L., Creteil L'Echat, France
CANBERRA ELEKTRONIK GmbH, Ottobrunn, West Germany
CANBERRA/POSITRONIKA B.V., Eindhoven-Woensel, Netherlands
CANBERRA-STOLZ A.G., Widen-Mutschellen, Switzerland

Circle No. 29 on Reader Service Card



An argument for the existence of elementary weakly coupled scalar fields is illustrated by this diagram. The areas represent different symmetries, as explained in the text. Figure 3

cause the u and d quark masses are equal as because they are small. The masses of the hadrons are close to what they would be if chiral SU(2) ⊗ SU(2) were exact, and although this group is spontaneously broken, its isospin subgroup is not.)

No one yet knows *why* the quark masses should have this pattern, so it can not be said that the gauge theories explain the approximate symmetries of the strong interactions. However, the gauge theory of strong interactions does rule out a vast number of other conceivable symmetry patterns, those based on orthogonal, symplectic or exceptional Lie groups. Also, there must be just *one* isotopic doublet of quarks, with all other quarks singlets, for otherwise the isotopic spin group would not be SU(2) but something larger.

I should refer here to a problem that arises in this line of reasoning. With the pattern of quark masses described above, the approximate symmetry group of the strong interactions would not be SU(3) ⊗ SU(3), but SU(3) ⊗ SU(3) ⊗ U(1). The extra U(1) symmetry would have to be spontaneously broken, but this would give rise to a very light ninth pseudoscalar meson of mass less than $\sqrt{3}$ pion masses, which is not observed. This problem appears to have now been solved by the discovery by Gerard 't Hooft of non-perturbative effects, associated with the "instanton" solution of gauge theories, which eliminate the extra U(1) symmetry. With the group reduced to SU(3) ⊗ SU(3), the mass of the eta meson is just given by the Gell-Mann–Okubo value, $(\frac{4}{3} m_K^2 - m_\pi^2)^{1/2}$, in agreement with the observed value. There still is a possible problem with the partial width for the reaction

$$\eta \rightarrow \pi^+ + \pi^- + \pi^0$$

which is now predicted to be 65 eV. The two measurements of the η width gave partial widths of 630 ± 140 eV in 1967 and 204 ± 22 eV in 1974. This is a difficult experiment to perform and analyze, but another effort appears to be called for. Unfortunately, the discovery of the instanton raises questions about the point made above, that the strong interactions *must* conserve P and T. The issue is not settled.

Vector and axial-vector currents The effective weak interactions at low energy must take the form of a product of vector and/or axial-vector currents in any gauge theory.

Conserved vector currents In renormalizable gauge theories, the strangeness-conserving vector part of the hadronic charged weak current can not be anything but the operator $\bar{u}\gamma^\mu d$. (Here u and d are the Dirac fields of the quarks, and γ_μ is a Dirac matrix.) This is just the charged part of the conserved current of isotopic spin, as in the Feynman–Gell-Mann theory of the vector current.

Partially conserved axial-vector currents Similarly, the strangeness-conserving axial-vector part of the hadronic charged weak current can not be anything but the operator $\bar{u}\gamma^\mu\gamma_5 d$, which is one of the approximately conserved currents of chiral SU(2) ⊗ SU(2). This justifies the hypotheses used in derivations of the "current algebra" results for weak interactions, such as the Goldberger–Treiman and Adler–Weisberger relations.

First-class currents In the absence of strongly interacting scalar fields, the strangeness-conserving part of the hadronic charged weak currents must have charge-conjugation properties of the so-called "first-class" type only. The experimental situation here is not clear.

Weak radiative corrections In unified gauge theories, the emission and reabsorption of virtual intermediate vector bosons $W^\pm$ and Z^0 produces "radiative" corrections of the order of the fine-structure constant α. As long as there are no strongly interacting scalar fields, these order-α corrections can only affect the mass matrix of the quarks, and therefore can not violate C,P,T,S, etc, conservation. However, they may produce an important part of the mass difference between the u and d quarks. This quark-mass difference may be responsible for the non-electromagnetic violation of isotopic spin conservation, which has long been thought to be needed to explain the $\eta \rightarrow 3\pi$ decay rate as well as the fact that the neutron is heavier than the proton.

Universality The weak and electromagnetic interactions of any quark or lepton depend only on how it transforms under the SU(2) ⊗ U(1) gauge group. With all quarks and leptons in doublets, they all have essentially the same weak interactions. The only qualification here is that the upper and lower members of the quark doublets are in general different mixtures of quark fields of definite mass, so that we have universality of the Cabibbo type, involving various mixing angles such as the Cabibbo angle. In "minimal" theories of leptons with zero neutrino mass no such angles can enter.

The immediate problem in our present "low-energy" studies of neutral currents and new particles is to identify the various types of leptons and quarks, and to decide how they are grouped into representations of the weak and electromagnetic gauge group. Beyond this lies the problem of explaining the pattern of masses of leptons and quarks. There are many ideas on why some of these masses should be nearly equal and why some of them should be nearly zero, but no one so far has been able to incorporate these ideas into a realistic model.

Medium energy

The fundamental problem that colors all our thinking about the medium energy range has to do with the nature of the mechanism for the spontaneous breakdown of the gauge symmetry of weak and electromagnetic interactions. For every broken symmetry, there must be a spinless particle, a "Goldstone boson." (This may either be a physical massless particle, or else serve as the helicity-zero part of a massive gauge vector boson.) There are two broad possibilities for spontaneous symmetry breaking, corresponding to two different ideas as to the nature of these Goldstone bosons.

It may be that the Goldstone bosons are bound states, held together by some sort of strong force between quarks, leptons and other particles; this is called a *dynamical symmetry breakdown*. Alternatively, the Goldstone bosons may be elementary and associated with a weakly coupled multiplet of scalar fields in the Lagrangian, the vacuum expectation values of which break the symmetry. To be colorful, I will call this a *non-dynamical symmetry breakdown*.

The most familiar test of the nature of the spontaneous symmetry-breaking mechanism relates to the masses of the intermediate vector bosons. If the symmetry breaking is non-dynamical, and if the scalar fields form any number of weakly coupled doublets, then the ratio of the mass of the charged intermediate vector boson m_W to that of the neutral intermediate boson m_Z is $\cos\theta$, where θ is the "mixing angle" measured in *ratios* of neutral-current cross sections. This relation sets the scale of the strength of neutral-current weak interactions, and appears to be in good agreement with experiment. With θ between 33° and 40°, we expect m_W to be between 58 and 68 GeV, and m_Z between 75 and 82 GeV. If the symmetry breaking were dynamical in nature, there would be no reason to expect any specific relation between the intermediate vector boson masses. It will of course be tremendously exciting to learn directly what the masses of the intermediate vector bosons actually are.

The same experiments that seek out the W and the Z are likely to encounter a number of other particles, of zero spin, with properties that depend critically on the nature of the symmetry-breaking mechanism. If the symmetry breaking is dynamical, then in the absence of any elementary scalar fields a theory involving N quark flavors will automatically have an approximate $SU(N) \otimes SU(N)$ symmetry, consisting of all unitary transformations on the right- and left-handed quark fields. These are represented by the large ellipse in figure 3. This "accidental" symmetry is weakly broken by the weak and electromagnetic interactions themselves, and strongly spontaneously broken by whatever mechanism breaks the gauge symmetry of the weak and electromagnetic interactions. For every one of the spontaneously broken accidental symmetries of $SU(N) \otimes SU(N)$, corresponding to points between the two ellipses in figure 3, there is a Goldstone boson. Those that correspond to generators of gauge symmetries (the interior of the small circle in the figure) are eliminated by the familiar "Higgs mechanism," turning into the helicity-zero parts of intermediate vector bosons.

However, the Goldstone bosons that correspond to spontaneously-broken non-gauge accidental symmetries are real physical particles. For broken symmetries that are also intrinsically violated by the weak and electromagnetic interactions (points outside the large circle in figure 3), these are "pseudo-Goldstone bosons," with masses comparable to m_Z or m_W. In most theories there are also some spontaneously broken symmetries, which are not violated by the weak and electromagnetic interactions, but are also not themselves gauge symmetries; to each of these (represented by the region between the two circles and the two ellipses in figure 3) there corresponds a *true* Goldstone boson, a weakly coupled physical particle of zero mass and spin.

For instance, in the "standard" $SU(2) \otimes U(1)$ model with only left-handed currents, the weak and electromagnetic interactions are automatically invariant under $SU(2)$ transformations among the right-handed parts of the d and s quarks, which are not gauge transformations. This symmetry is certainly not realized in physical states; so with no scalar fields available to break it, it would have to be spontaneously broken, leading to a triplet of neutral massless Goldstone bosons coupled to s and d quarks. These bosons would certainly have been seen, for instance in the decay of a K meson into a pi meson and a Goldstone boson, so one can rule out this kind of theory. Similar remarks apply in other models and furnish what I think is the strongest empirical argument for the existence of weakly coupled elementary scalar fields and a non-dynamical symmetry breaking.

If indeed there are elementary scalar fields the vacuum expectation values of which break the $SU(2) \otimes U(1)$ gauge symmetry, and if the theory is renormalizable, then some of these fields must correspond to massive scalar particles, which are not Goldstone bosons and are not killed off by the Higgs mechanism. These have come to be called *Higgs bosons*.

Too massive to be seen

The properties of the Higgs bosons have been under intensive theoretical study lately. One can understand their properties qualitatively, by returning to the simple sort of model used to study spontaneous symmetry breaking from the early 1960's on. If the Hamiltonian of the scalar fields has a bare-mass term $-\frac{1}{2}\mu^2\phi^2$ and a self-interaction term $\frac{1}{4}f\phi^4$ (where ϕ^2 is a sum of squares of scalar field com-

Glossary

Asymptotic freedom The property of some gauge theories of the strong interactions, that the strong interactions grow steadily weaker at high energy.

Chiral Refers to groups of transformations that act differently on the left- and right-handed parts of fermion fields. The currents associated with chiral symmetry groups are of both vector and axial-vector type.

Color The quantity that determines the coupling of any particle to the "gluon" field in the gauge theory of strong interactions. Quarks of each flavor are supposed to come in three different colors.

First-class currents Weak-interaction currents with the same charge symmetry (or G-parity) properties as the currents in the Fermi theory of beta decay.

Flavor The label used to distinguish different types of leptons (ν_e, e, ν_μ, μ, etc), and different color triplets of quarks (u, d, c, s, etc). The interactions of intermediate vector bosons and photons with quarks and leptons depend on flavor, probably not on color.

Gauge theories Theories invariant under a group of internal symmetries the effect of which varies from point to point in space and time. Such theories necessarily involve vector fields, known as "gauge fields," like the electromagnetic potentials or the Yang–Mills fields.

Gluons The vector fields of the gauge theory of strong interactions.

Goldstone bosons Particles of zero mass and zero spin, which accompany the spontaneous breakdown of symmetries.

Higgs mechanism The feature of spontaneously broken gauge symmetries, that the Goldstone bosons are not physical particles, but instead provide the helicity-zero states of gauge vector bosons of non-zero mass. The remaining physical spin-zero particles are known as "Higgs bosons."

Intermediate vector bosons Massive spin-one particles that are supposed to mediate the weak interactions. In gauge theories of the weak and electromagnetic interactions, the intermediate vector bosons are the quanta of gauge fields, which are given masses by the Higgs mechanism.

Neutral currents Weak interactions in which no charge is exchanged between fermions. In unified gauge theories of the weak and electromagnetic interactions, the neutral-current interactions are produced by exchange of a neutral intermediate vector boson, the Z^0.

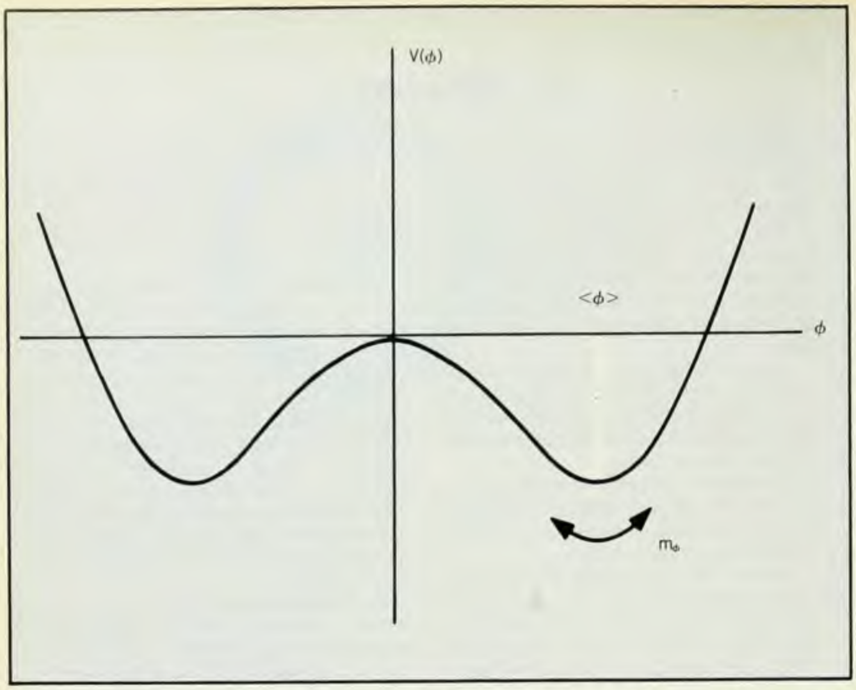
Pseudo-Goldstone bosons Goldstone bosons associated with the spontaneous breakdown of approximate accidental symmetries, as opposed to exact fundamental symmetries.

Quarks Hypothetical spin one-half particles, which are supposed to be the fundamental constituents of all hadrons.

Renormalizability The property of some quantum field theories, that all infinities can be absorbed into a renormalization of physical parameters such as mass and charge. Renormalizable theories involve only a finite number of free parameters.

SU(*N*) The group of unitary $N \times N$ matrices with unit determinant.

U(1) The group of complex numbers with modulus unity.



A "potential" for the determination of properties of the scalar field; it is of the quartic form required for renormalizability. The minimum locates the vacuum expectation value of the scalar field, and the curvature there gives the mass of the Higgs scalar boson. Figure 4

ponents) then the "potential" for this field takes the double-well shape shown in figure 4.

The vacuum expectation value $\langle\phi\rangle$ of the field ϕ is just its value $\mu/\sqrt{f}$ at the bottom of the well. The mass of the Higgs particle is the curvature $\mu\sqrt{2}$ of the well at this point, and can therefore be written as $\langle\phi\rangle\sqrt{2f}$. Now, we know $\langle\phi\rangle$ reasonably well, because the masses of the intermediate vector bosons in unified theories are of order $e\langle\phi\rangle$ (where e is the electronic charge), while the Fermi coupling constant G_F is of order $e^2/m_W{}^2$; hence $\langle\phi\rangle$ must be of order $G_F{}^{-1/2}$, or 300 GeV. We do not know the scalar self-coupling f, but it can hardly be less than of order e^4 because a coupling this strong wuld be induced by higher-order effects of virtual W particles. Hence, with f larger than the order e^4, the Higgs mass $\langle\phi\rangle\sqrt{2f}$ must be larger than of order $e^2G_F{}^{-1/2}$, which is a few GeV.

In fact, this argument can be made quite precise: In the simplest SU(2) ⊗ U(1) model, the lower bound turns out to be 4.9 GeV. On the other hand, we would like to think that the Higgs fields are weakly coupled, with f much less than 1, because otherwise our use of perturbation theory would be invalid, and we would not be able to calculate such quantities as m_Z/m_W. With $f \ll 1$, the Higgs mass must be much less than 300 GeV. The most common guess is that f is of order e^2, giving a Higgs mass of order $eG_F{}^{-1/2}$, just about the same as the intermediate vector bosons. No wonder they have not been seen!

We can also say something about the couplings of the Higgs bosons. If the scalar field ϕ has a coupling to quarks or leptons of the familiar Yukawa form $f'\phi\bar{\psi}\psi$, then the spontaneous symmetry breakdown will evidently give the fermion a mass of order $f'\langle\phi\rangle$ so the coupling of a Higgs boson to a lepton or a quark of mass m is expected to be of order $m/\langle\phi\rangle$, or m/300 GeV. The exchange of a Higgs particle of mass m_H between two fermions of mass m_F would then produce an effective Fermi interaction, diagrammed in figure 5, with coupling weaker than the usual Fermi coupling (300 GeV)$^{-2}$ by a factor of order $(m_F/m_H)^2$. Even if m_F is as large as 1 GeV and m_H is as small as 30 GeV, these effects are "milli-weak," down in strength by a factor of order 10^{-3} from the usual weak interactions. Thus we can easily understand not only why Higgs particles themselves have not been seen, but also why their effects have not been seen.

But effects of virtual Higgs bosons actually may have been seen, and long ago. I emphasized in the discussion of "low-energy" physics that the strong interactions automatically obey a number of symmetries (C,P,T,S, etc) whether or not these are assumed from the outset. There are similar accidental symmetries in sufficiently simple models of the weak interactions, *except for effects of the Higgs bosons.* Therefore effects of virtual Higgs bosons may be detected as small violations of an apparent symmetry of the weak interactions.

One example is CP nonconservation. A theorem of M. Kobayashi and K. Maskawa says that in a minimal SU(2) ⊗ U(1) theory (with only four quarks, the usual leptons and left-handed currents) the couplings of quarks and leptons to the photon, $W^\pm$ and Z^0 must conserve CP, whether or not CP conservation is assumed from the outset. The same is true of the couplings of the quarks and leptons to Higgs bosons, provided there is only one doublet of scalar fields. However, if there are arbitrary numbers of scalar fields and if we do not assume CP conservation as an *a priori* symmetry, then the couplings of Higgs bosons to quarks and leptons will violate CP strongly.

In such a theory, as first proposed by T. D. Lee, exchange of virtual Higgs bosons will produce CP-violating processes such as the reaction $K_L{}^0 \to 2\pi$, the Feynman diagram of which is shown in figure 6. As already mentioned, Higgs exchange tends to be a milliweak effect, so we can understand in a natural way why the CP-violating phase of the $K_L{}^0 \to 2\pi$ amplitude is so small. Higgs exchange would also produce a neutron electric dipole moment of order 10^{-24} e cm, just about at the level of the present empirical upper limit. It will be interesting to see whether a non-vanishing electric dipole moment turns up with the next round of improvements in this experiment.

Muon number—not conserved?

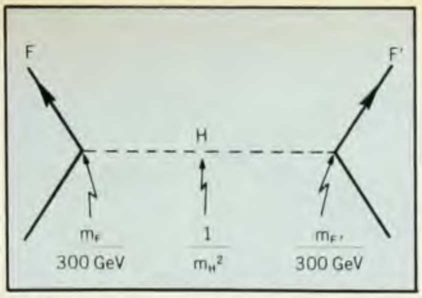
Another possible example of an approximate accidental symmetry of the weak interactions is muon conservation. In the simplest SU(2) ⊗ U(1) model with two doublets of left-handed leptons $(\nu, e^-)_L$ and $(\nu', \mu^-)_L$ and two right-handed leptons $e_R{}^-$, $\mu_R{}^-$, we can always define the electronic and muonic leptons so that muon number is conserved by the ordinary weak interactions. Muon number will also be conserved by Higgs-exchange effects, if there is only one doublet of elementary scalar fields in the theory. However, if there are arbitrary numbers of scalar fields, and if we do not assume muon conservation as an *a priori* symmetry principle, then virtual Higgs bosons will produce muon-nonconserving processes such as $\mu \to e + \gamma$. Oddly enough, because of the smallness of the lepton masses it turns out that the dominant contribution comes from *two*-loop diagrams such as the one shown in figure 7. James Bjorken and I estimate that these diagrams give a branching ratio for $\mu \to e + \gamma$ of order $(\alpha/\pi)^3 \approx 10^{-8}$, or somewhat less if the mixing of the different Higgs bosons is not maximal.

An experiment to look for the decay of a muon into an electron and a photon is now under way at the Swiss Institute for Nuclear Research. There have been rumors of a certain number of events that may be decays of this type, corresponding to a branching ratio of order 10^{-9}. Whatever the outcome of this particular experiment, it is clear that experiments that could measure the branching ratio for $\mu \to e + \gamma$ or set a new upper bound at the level of 10^{-10} or so would be of the highest importance. Another process that appears to be ripe for renewed experimental study is the conversion of muons into electrons in muonic atoms.

I certainly would not argue that Higgs-boson exchange is the only possible source of CP or muon-number nonconservation. There are several models with extra leptons, in which muon nonconservation occurs through W exchange. One of those models, proposed by T. P. Cheng and L.-F. Li (and somewhat modified by Bjorken, Kenneth Lane and me, as well as by Benjamin Lee and Robert Schrock) will become particularly attractive if the preliminary indications of parity conservation in atoms are confirmed. There are also a number of models with extra quarks (recently analyzed in great generality by Lee) in which CP is violated by W exchange. In such theories the electric dipole moment of the neutron is expected to be very small, about 10^{-30} e cm, and the overall milliweakness or superweakness of CP violation is unexplained. To complicate matters further, we can consider theories with extra quarks and leptons and more than one multiplet of scalar fields—in this case, either CP violation or muon nonconservation could occur through *both* Higgs exchange and intermediate-vector-boson exchange.

However all this turns out, there is one moral that I would like to draw from these speculations about the nature of approximate symmetries. There are some symmetries, such as the $SU(2) \otimes U(1) \otimes SU(3)$ gauge symmetries, which appear to be exact (whether or not they are spontaneously broken). We really do not have any good ideas on why Nature should respect these particular symmetries, and, for the moment, we have to take them as axiomatic. On the other hand there are many symmetries that are not exactly obeyed. It is inconceivable that an approximate symmetry could be anything really fundamental, so we are led to seek dynamical explanations. Remarkably, as we have seen, it *is* possible to explain why P, C, T, S, etc must be conserved by the strong and electromagnetic interactions in a renormalizable gauge theory. We may also be able to explain, depending on the menu of quarks and leptons, why CP and muon number are approximately conserved even by the weak interactions.

This can be turned around—if we find dynamical arguments showing that any possible violation of some supposed symmetry would automatically be too weak to have surfaced in any existing experiment, then we may begin to suspect that the symmetry is only a dynamical accident, and that violations will eventually turn up. This change in point of view seems to be occurring now for muon number. Baryon number may be next.



Exchange of Higgs scalars. The effective coupling between fermions F and F′ that is induced by the exchange of a Higgs boson H is the product of the three factors indicated on the Feynman diagram; it is therefore weaker, by a factor $m_F m_{F'}/m_H{}^2$, than the usual Fermi coupling of $(300\text{ GeV})^{-2}$. Figure 5

No one thinks that a non-simple gauge group such as $SU(3) \otimes SU(2) \otimes U(1)$ could be the end of the story. From the beginning of the current wave of interest in gauge theories, it has been speculated that the weak and electromagnetic gauge group might be only a part of a larger simple gauge group $\mathcal{G}$. (By "simple" I mean here that the symmetry allows only a single independent gauge coupling constant. This is almost the same as the group-theoretic definition.)

High energy

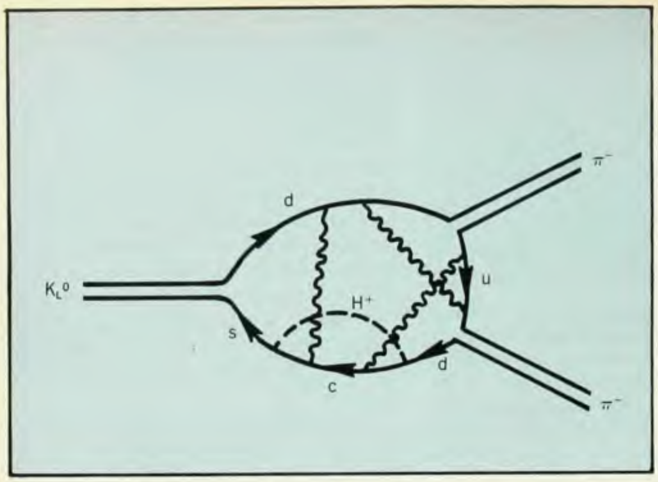
This would not only determine a unique value for the mixing angle θ, and satisfy our sense of tidiness; the incorporation of the U(1) subgroup into a larger simple gauge group would also account for the observed fact that electric charge is quantized. To explain why we do not see effects of the other gauge vector bosons of the big simple group $\mathcal{G}$ it would be necessary to assume that they are much heavier than the W and Z, which means that the $SU(2) \otimes U(1)$ subgroup is not as strongly spontaneously broken as the rest of $\mathcal{G}$.

The discovery of asymptotic freedom opened up the possibility of including strong interactions in this picture. We can suppose that the over-all simple group $\mathcal{G}$ suffers a superstrong spontaneous breakdown to $SU(3) \otimes SU(2) \otimes U(1)$, so that all gauge vector bosons of $\mathcal{G}$ other than the $W^{\pm}$, Z^0, the photon and gluons pick up enormous masses, characterized by some mass scale M much larger than 300 GeV. If measured at the really high energies of order M, the gauge coupling constants of SU(3), SU(2) and U(1) would all be found to be equal, aside from group-theoretic factors of order one, which depend on how these subgroups are embedded in $\mathcal{G}$. However, asymptotic freedom means that the SU(3) coupling constant increases with decreasing energy, so at sufficiently low energies it can become strong, of order unity, while the SU(2) and U(1) couplings are still of order e. From this point of view, the strong interactions *must* be asymptotically free; otherwise they would not be strong. (The ratio of the SU(2) and U(1) coupling constants would be strongly affected by such renormalization effects, so that the weak mixing angle would be very different from the group-theoretic value that would be observed at energies of order M.) Over most of the range from "high" energies of order M down to "low" energies of a few GeV, the SU(3) coupling constant will still be fairly small, so its growth with decreasing energy will only be logarithmic. Hence M must be truly enormous to give the strong interactions room to grow strong. Howard Georgi, Helen Quinn and I estimated in one case that $M \approx 10^{17}$ GeV.

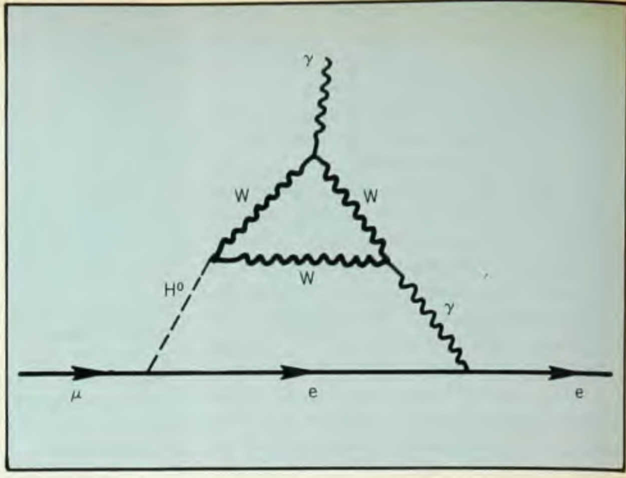
Whatever the actual value of M, we are forced to consider at least two very different levels of spontaneous symmetry breaking: a "superstrong" breakdown of $\mathcal{G}$ into $SU(3) \otimes SU(2) \otimes U(1)$ with superlarge intermediate vector boson masses, and the "ordinary" breakdown of $SU(2) \otimes U(1)$ into the electromagnetic U(1) gauge invariance, with $W^{\pm}$ and Z^0 masses in the medium energy range.

We live among the debris of the superstrong symmetry breakdown. The particles we observe at "low" and "medium" energies are just those that somehow escaped getting a superlarge mass of order M from the superstrong breakdown of the original gauge symmetry group $\mathcal{G}$. The superstrong symmetry breakdown presumably does not give such particles any mass at all. This is of course true for the gauge vector bosons of the non-superstrongly-broken subgroup $SU(3) \otimes SU(2) \otimes U(1)$. It is also true of any fermions that are kept massless by any non-superstrongly-broken chiral symmetries in $\mathcal{G}$—these must include all the observed leptons and quarks. But what about the scalar bosons? If they all pick up superlarge masses of order M (or become Goldstone bosons), none will be available to produce the "ordinary" breakdown of $SU(2) \otimes U(1)$ at medium energies. How can any of the scalar bosons (aside from Goldstone bosons) be left massless by the superstrong symmetry breakdown?

As far as I know, the only natural way to keep a scalar boson massless is to have a "supersymmetry," a symmetry of the Wess–Zumino type, which puts scalar fields in the same multiplet as massless fermion fields. There is no supersymmetry evident in the table of particle masses, so we would have to suppose that whatever supersymmetry survived the superstrong symmetry breakdown was spontaneously broken along with $SU(2) \otimes U(1)$ at medium energies. The trouble with this idea is that supersymmetries are notoriously hard to break. The vacuum expectation values of the scalar fields in any theory are just the coordinates of the point at which a "potential" function like that shown in figure 4 takes its minimum value. In supersymmetric theories the potential is positive, and can take the value zero only for field values invariant



Higgs exchange as a source of the CP-violating process $K_L^0 \to \pi^+ + \pi^-$. The solid lines represent quarks; the wavy lines, typical virtual gluons, and the dashed line, a charged Higgs boson. Figure 6



A typical two-loop contribution to muon decay into an electron and a photon. The solid lines here are leptons, the wavy lines photons and the dashed line is a neutral Higgs boson. Figure 7

under the supersymmetry.

If there is such a supersymmetric set of scalar field values, then it clearly is at the minimum of the potential, and the supersymmetry can not be broken spontaneously. In some special cases, found by P. Fayet and John Iliopoulos, there is no supersymmetric set of scalar field values, and the supersymmetry *must* be broken, even in the tree approximation. However, I do not know of any way in which a supersymmetry that escapes being spontaneously broken at the superstrong level could ever be broken at all. Any progress towards a solution would be valuable.

Towards medium-energy experiments

These issues will become of really pressing concern when experiments begin to explore the "medium" energy range. If the quantum field theory that describes physical phenomena at low and medium energies involves just those particles that were left massless by the superstrong symmetry breakdown, then it should be governed by a scale-invariant Lagrangian. The "ordinary" spontaneous symmetry breakdown in such a theory must be of the Coleman–E. Weinberg type, produced by a balance between quartic self-couplings and "radiative" corrections.

Scale invariance is itself spontaneously broken in this way, and one of the Higgs bosons must be the corresponding Goldstone boson, known as a "dilaton" or "scalon." But scale invariance is also broken by the renormalization procedure, so the scalon is a pseudo-Goldstone boson, with a mass of order e^2 times 300 GeV. In an SU(2) ⊗ U(1) theory in which there are no other particles as heavy as the $W^\pm$ and Z^0, the mass of the scalon was found by Eldad Gildener and me (for $\theta = 35°$) to be just 7.3 GeV. Adding Higgs bosons with masses of order m_W raises the scalon mass; adding fermions this heavy lowers it. The discovery of the scalon may be one of the most enlightening products of medium-energy research.

It is also possible that there are *no* scalar bosons that escape getting masses of order M from the superstrong symmetry breakdown. In this case, the ordinary breakdown of SU(2) ⊗ U(1) would have to be dynamical. We have already seen various difficulties with this idea and now we encounter another: A dynamical spontaneous breakdown of SU(2) ⊗ U(1) would require that there is some sort of interaction strong enough at energies of order 300 GeV to produce Goldstone bosons as bound states. None of the known gauge couplings is this strong. Indeed, the ordinary color-SU(3) strong interactions presumably do produce the dynamical breakdown of chiral SU(2) ⊗ SU(2), which is supposed to be responsible for the appearance of the pion as a Goldstone boson, but this is a "low"-energy phenomenon, as shown by the value $F_\pi \approx 190$ MeV of the fundamental scale factor of the soft-pion theorems. These strong interactions are already so weak at energies of a few GeV that they can be treated by perturbation theory, and at 300 GeV they are even weaker.

One way out of this difficulty is to suppose that color SU(3) is only a subgroup of a larger strong-interaction gauge group $\mathcal{G}_s$, and that the subgroup of $\mathcal{G}$ which survives the superstrong symmetry breakdown is not SU(3) ⊗ SU(2) ⊗ U(1), but $\mathcal{G}_s$ ⊗ SU(2) ⊗ U(1). (The idea of Jogish Pati and Abdus Salam that lepton number is a "fourth color" suggests that $\mathcal{G}_s$ might be SU(4).) We could then suppose that the $\mathcal{G}_s$ gauge coupling does become strong when the energy drops to 300 GeV, and that it is this force that binds the Goldstone bosons needed to break SU(2) ⊗ U(1) down to U(1) *and* $\mathcal{G}_s$ down to color SU(3). (We do not know how to calculate the color SU(3) gauge coupling that would result, but it might well be a little weaker than the $\mathcal{G}_s$ couplings at 300 GeV.) With $\mathcal{G}_s$ a larger group than SU(3), the decrease of the $\mathcal{G}_s$ gauge coupling with increasing energies above 300 GeV is likely to be faster than for SU(3), so the superlarge mass M at which all the gauge couplings become comparable could be very much less than 10^{17} GeV. The real test of this class of theories would be to carry experiments to center-of-mass energies of order 300 GeV and see if new strong interactions set in, which act on leptons as well as hadrons.

Back to gravitation

If the superstrong symmetry breakdown really does involve masses as large as 10^{17} GeV it can not be properly understood without a satisfactory quantum theory of gravitation. The problem here is one of renormalizability. There has been exciting progress lately in building supersymmetric theories in which the graviton appears in multiplets with other particles, but, as far as I know, there is no demonstration that any of these theories are renormalizable. Indeed, it is not clear that renormalizability is even the correct constraint to impose here.

In any case, I find it very satisfying that the study of quantum field theory has led us inexorably back to what was the first successful field theory, the theory of gravitation.

Suggested readings

For reviews of the subject, see

- Rapporteur's talks at recent International Conferences on High Energy Physics, by B. W. Lee in 1972, J. Iliopoulos in 1974 and A. Slavnov in 1976.
- S. Weinberg, Rev. Mod. Phys. **46**, 255 (1974).
- J. C. Taylor, *Gauge Theories of the Weak Interactions*, Cambridge U. P. (1976).

For a brief discussion of cosmological implications of gauge theories, see

- S. Weinberg, *The First Three Minutes—A Modern View of the Origin of the Universe*, Basic Books, New York (1977). □