Annu. Rev. Nucl. Part. Sci. 1993. 43:209–52

EFFECTIVE FIELD THEORY[1]

Howard Georgi

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

KEY WORDS: divergences, renormalization, regularization, Appelquist-Carrazone

---

CONTENTS

Prologue

When I agreed to write this review article (in a moment of weakness) in 1991, I didn't fully realize how poorly suited I am to this kind of activity. As I gathered ideas and materials for the article over the next year, it dawned on me that I hadn't even read a review article in 25

[1] Research supported in part by the National Science Foundation under Grant PHY-8714654, and in part by the Texas National Research Laboratory Commission, under Grant RGFY9206.

0163-8998/93/1201–0209$02.00

years. I'm not quite sure of their purpose. Obviously, I cannot summarize all the work that has gone on in the field in any detail. I could try to give a guide to the literature, but it is easy to summarize the proper approach to the old literature on this (or almost any) subject. Two words suffice. Ignore it! With rare exceptions, old papers are difficult to read because the issues have changed over the years.

The model that I use instead is my "textbook" on weak interactions (1), which was originally intended as a thinly disguised monograph on effective field theory. There are more words than formulas in this review. I concentrate on what I think are the critical ideas from which all the applications flow, illustrating the technical issues in just a few important and rather new examples. I refer to many of the old papers on the subject in Section 2, but usually in the hope that people will not actually go back and read them.

I introduce the effective field theory idea in Section 1. Section 2 contains a brief review of some of the history of effective field theory. In Section 3, I discuss the process of matching from one effective field theory to the next in some detail. Finally, in Section 4, I discuss a few applications of the effective field theory idea. These are illustrative, but not at all exhaustive. For example, I do not mention the beautiful example of heavy quark effective theory (2) at all because it has previously been reviewed in this series (2a). I end this review with a frankly speculative discussion of nonperturbative matching corrections.

1. EFFECTIVE FIELD THEORY

One of the most astonishing things about the world in which we live is that there seems to be interesting physics at all scales. Whenever we look in a previously unexplored regime of distance, time, or energy, we find new physical phenomena. From the age of universe, about 10^{18} sec, to the lifetime of a W or Z, a few times 10^{-25} sec, in almost every regime we can identify physical phenomena worthy of study.

To do physics amid this remarkable richness, it is convenient to be able to isolate a set of phenomena from all the rest, so that we can describe it without having to understand everything. Fortunately, this is often possible. We can divide the parameter space of the world into different regions, in each of which there is a different appropriate description of the important physics. Such an appropriate description of the important physics is an "effective theory." The two key words here are *appropriate* and *important*. The word *important* is key because the physical processes that are relevant differ from one place in parameter space to another. The word *appropriate* is key because there is no single description of physics that is useful everywhere in parameter space.

The common idea is this: if there are parameters that are very large or very small compared to the physical quantities (with the same dimension) we are interested in, we may get a simpler approximate description of the physics by setting the small parameters to zero and the large parameters to infinity. Then the finite effects of the parameters can be included as small perturbations about this simple approximate starting point.

This is an old trick, without which much of our current understanding of physics would have been impossible. We use it without thinking about it. For example, we still teach Newtonian mechanics as a separate discipline, not as the limit of relativistic mechanics for small velocities. In the (familiar) region of parameter space in which all velocities are much smaller than the speed of light, we can ignore relativity altogether. It is not that there is anything wrong with treating mechanics in a fully relativistic fashion. It is simply easier not to include relativity if you don't have to.

This simple example is typical. It is not necessary to use an effective theory, if you think that you know the full theory of everything. You can always compute anything in the full theory if you are sufficiently clever. It is, however, very convenient to use the effective theory. It makes calculations easier, because you are forced to concentrate on the important physics.

In the particle physics application of effective theories, the relevant parameter is distance scale. In the extreme relativistic and quantum mechanical limit of interest in particle physics, this is the only relevant parameter. The strategy is to take any features of the physics that are small compared to the distance scale of interest and shrink them down to zero size. This gives a useful and simple picture of the important physics. The finite size effects that you have ignored are small and can be included as perturbations.

Again, this process is very familiar. We use it, for example, in the multipole expansion in electrodynamics, or in replacing a physical dielectric with a uniform one. However, in a relativistic, quantum mechanical theory, in which particles are created and destroyed, the construction of an effective theory (now an effective quantum field theory—EQFT) is particularly interesting and useful. An EQFT is particularly useful because among the short-distance features that can be ignored in an effective theory are all the particles too heavy to be produced. Eliminating heavy particles from the effective theory produces an enormous simplification. An EQFT is particularly interesting because of the necessity of ultraviolet regularization. This makes the process of constructing the effective theory nontrivial because the limit in which the small-distance scales are taken to zero must be handled carefully. One consequence of the ultraviolet behavior is the renormalization group running of coupling constants with the renormalization scale, μ. Going to the effective theory actually changes the running of coupling constants by trading logarithmic dependence on heavy-particle masses for scale dependence. I have much more to say about this issue below.

The result of eliminating heavy particles is inevitably a nonrenormalizable theory, in which the nontrivial effects of the heavy particles appear in interactions with dimension higher than four. In the full theory, these effects are included in the nonlocal interactions obtained by "integrating out" the heavy particles. These interactions, because of their nonlocal nature, get cut off for energies large compared to the heavy-particle masses. However, in the effective theory, we replace the nonlocal interactions from virtual heavy-particle exchange with a set of local interactions, constructed to give the same physics at low energies. In the process, we have modified the high energy behavior of the theory, so that the effective theory is only a valid description of the physics at energies below the masses of the heavy particles. Thus the domain of utility of an effective theory is necessarily bounded from above in energy scale.

Likewise, at least if the effective theory itself describes light particles with nonzero mass, the domain of utility of the effective theory is bounded from below. At sufficiently small energy scales, below the masses of the heaviest particles in the effective theory, it is both possible and useful to change theories yet again to a new effective theory from which the heaviest particles have been removed. While the upper bound is an absolute, this lower bound is simply a convenience.

In the extreme version of the effective-field-theory language, we can associate each particle mass with a boundary between two effective theories. For momenta less than the particle mass, the corresponding field is omitted from the effective theory. For larger momenta, the field is included. The connection between the parameters in the effective theories on either side of the boundary is now rather obvious. We must relate them so that the description of the physics just below the boundary (where no heavy particles can be produced) is the same in the two effective theories. In lowest order, this condition is simply that the coupling constants for the interactions involving the light fields are continuous across the boundary. Heavy-particle exchange and loop effects introduce corrections as well as new nonrenormalizable interactions, which are discussed in detail below. The relations between the couplings imposed by the requirement that the two effective theories describe the same physics are called "matching conditions." The matching conditions are evaluated with the renormalization scale μ in both theories of the order of the boundary mass to eliminate large logarithms.

The process of matching is sometimes thought of as a two-step process. In the first step, you integrate out the heavy particles and produce nonlocal interactions among the light particles. In the second step, you perform an operator product expansion on the nonlocal interactions to produce the local interactions in the effective theory. In fact, however, this two-step picture is not quite right. I show where it breaks down in Section 3.2 below. But even where the picture is valid, it is not really very useful. The reason is that the process of doing the operator product expansion is not very different from the process of computing the matching corrections. You might as well do the matching in a single step. We see how this works in detail below.

If we had a complete renormalizable theory at high energy, we could work our way down to the effective theory appropriate at any lower energy in a systematic way. Beginning with the mass M of the heaviest particles in the theory, we could set $\mu = M$ and calculate the matching conditions for the parameters describing the effective theory with the heaviest particles omitted. Then we could use the renormalization group to scale μ down to the mass M' of the next heaviest particles. Then we would match onto the next effective theory with these particles omitted and then use the renormalization group again to scale μ down further, and so on. In this way, we obtain a descending sequence of effective theories, each one with fewer fields and more small renormalizable interactions than the last. This "top down" approach to effective field theory is a very convenient way of organizing field theory calculations. We discuss several examples of this procedure in the sections to come.

1.1 *The Principles of Effective Field Theory*

Effective field theory is more than a convenience. There is another way of looking at it, however, that corresponds more closely to what we actually do in physics. We can look at this sequence of effective theories from the "bottom up." In this view, we do not know what the renormalizable theory at high energy is, or even that it exists at all. Having seen it several times in the last 50 years, we are now used to the idea that there are important interactions at many different energy scales, some of them probably so large that we cannot see them directly. Certainly not now. Perhaps not ever. Nevertheless, we can use an effective field theory to describe physics at a given energy scale, E, to a given accuracy, ϵ, in terms of a quantum field theory with a finite set of parameters. We can formulate the effective field theory without any reference to what goes on at arbitrarily small distances.

The "finite set of parameters" part sounds like old-fashioned renormalizability. However, the dependence on the energy scale, E, and the accuracy, ϵ, is the new feature of effective field theory. It arises because we cannot possibly know, in principle, what is going on at arbitrarily high energies. However, we can parameterize our ignorance in a useful way. The effect of physics at high energy on the physics at the scale, E, can be described by a tower of interactions, with integral mass dimension from two to infinity, beginning with conventional renormalizable interactions but going on to include nonrenormalizable interactions of arbitrarily high dimension.

The principles that govern the tower of interactions are these:

1. There are a finite number of parameters that describe the interactions of each dimension, $k - 4$.
2. The coefficients of each of the interaction terms of dimension $k - 4$ is less than or of the order of

$$\frac{1}{M^k}, \tag{1.}$$

where $E < M$ for some mass M, independent of k.

These two conditions are the principles of effective field theory. They ensure, at least in perturbation theory, that only a finite number of parameters are required to calculate physical quantities at an energy E to an accuracy ϵ, because the contribution of interactions of dimension k is proportional to

$$\left(\frac{E}{M}\right)^k. \tag{2.}$$

Thus we need only include terms up to dimension k_ϵ

$$\left(\frac{E}{M}\right)^{k_\epsilon} \approx \epsilon \rightarrow k_\epsilon \approx \frac{\ln(1/\epsilon)}{\ln(M/E)}, \tag{3.}$$

of which there are only a finite number.

Of course, as you go up in energy, the nonrenormalizable interactions for any fixed k become more important, and k_ϵ increases. This is a signal that you are getting close to new physics. Before you reach energies of order M, the nonrenormalizable interactions disappear and are revealed as renormalizable, or at least less nonrenormalizable, interactions with a still higher scale, M'. Then you have a new effective theory, and you can start the process over again.

The philosophical question underlying old-fashioned renormalizability is this: How does this process end?

It is possible, I suppose, that at some very large energy scale, all the nonrenormalizable interactions disappear, and the theory is simply renormalizable in the old sense. This seems unlikely, given the difficulty with gravity.

It is possible that the rules change dramatically, as in string theory.

It may even be possible that there is no end, simply more and more scales as one goes to higher and higher energy.

Who knows?

Who cares?

In addition to being a great convenience, effective field theory allows us to ask all the really scientific questions that we want to ask without committing ourselves to a picture of what happens at arbitrarily high energy.

Renormalizability has been replaced by the constraints (Equations 1 and 2) on the tower of interactions in the effective theory.

1.2 *Wilson versus Continuum EFT*

Within the general framework of the effective field theory idea, there are two rather different approaches, which I call the Wilson approach and the continuum effective field theory approach. It is the second of these that I discuss in detail here, but I begin by explaining why I think that they are different. I argue that the two take a very different approach to renormalization.

In Wilson effective theory, the fundamental question is "How does the full theory change as you integrate out high momentum modes and look at it at larger distances?" This question fits in nicely with a physical renormalization scheme such as momentum space subtraction and physical renormalization.

In what I call continuum effective field theory, the question is "How do we modify the theory to allow the use of a mass-independent scheme and still get the physics right?" The idea is to put in by hand as much as possible of the dependence on distance scale. The more of the physics of distance scale that is put in by hand, the easier it becomes to extract the physics that you really care about.

1.3 *Why Continuum Effective Field Theory?*

One of my motivations in agreeing to write this article was a question that my colleague Sidney Coleman asked me: What's wrong with form factors? What's wrong with just integrating out heavy particles and large momentum modes, à la the Wilson approach, and using the resulting nonlocal theory as your interaction?

The answer of course is "Nothing"—but this is not an effective field theory calculation. It is just a way of doing the full theory calculation. But the fact that one of the world's greatest field-theorists would ask such a question convinced me that the idea of continuum effective field theory is not universally understood.

The real answer to the question "Why continuum effective field theory?" is "Because it is easier!" The advantages are as follows.

1. Concentration on relevant physics: It allows us to deal just with the particles that we actually know about and the interactions that we already see, and to postpone speculation about higher energies.
2. Consistency with a mass-independent renormalization scheme: It allows us to use a convenient scheme like $\overline{\mathrm{MS}}$ and still get the physics right. I say much more about this below.
3. Dealing efficiently with infrared divergences: As I show below, the effective theory calculations can be organized explicitly to avoid infrared divergences. In particular, calculations of matching corrections are automatically infrared finite.

Of course, the real answer to Coleman's question is that it is the wrong question. The right question is "How is just integrating out heavy particles and using the resulting nonlocal theory as your interaction different from a real effective field theory calculation?" It is this difference that I think is not as widely appreciated as it should be. This is what I want to describe in some detail.

2. ROOTS

First, however, I trace a few of the historical roots of the modern view of continuum effective field theory. Of course, Ken Wilson's work in the 1960s lies at the root of much of the effective field theory idea. I do not discuss it here, because I want to concentrate on the developments that were uniquely related to the continuum theory.

2.1 *Chiral Lagrangians*

Chiral Lagrangian techniques were developed by Weinberg, Dashen, and others in the late 1960s as a short-cut to current-algebra derivations

(3). The struggle to extricate chiral Lagrangians from current algebra (see 4) and to make sense of loop calculations in these theories was one of the important spurs to the development of the effective field theory machinery (3–5; note that the logarithmic effects discussed in these works would, in modern parlance, be described in terms of renormalization group running). The chiral Lagrangian has also grown, at the hands of Gasser & Leutwyler, into the most nontrivial and important example of the use of effective field theory (6). In modern language, the point is simply that there exists a limit of QCD in which the pions, kaons, and etas are massless Goldstone bosons of spontaneously broken SU(3) $\times$ SU(3) chiral symmetry. In this limit, and at low momenta, the properties of these particles should be severely constrained by the chiral symmetry. The simplest and most systematic way to impose these constraints is to build a chiral Lagrangian describing only the Goldstone bosons, but incorporating the full chiral symmetry of QCD, nonlinearly realized (7). A chiral Lagrangian based on a nonlinear representation for SU(2) $\times$ SU(2) was constructed by Weinberg (7a).

If the chiral symmetries of QCD were exact, we could extract arbitrarily precise predictions from the chiral Lagrangian. The chiral Lagrangian can be organized in powers of the Goldstone boson momenta, p. The leading term, of order p^2, depends on only a single parameter, f_π. At very low energies, all Goldstone boson scattering processes are determined by this one parameter, up to corrections of the order of p^2/Λ^2, where $\Lambda \approx 1$ GeV is a parameter that I call the chiral symmetry-breaking scale and that measures the convergence of the momentum expansion.

In the real world, where the chiral symmetries are broken by the quark masses and by the electromagnetic interactions, we cannot get rid of the higher terms in the chiral Lagrangian by going to arbitrarily low energy in physical processes, because the particles are stuck on their mass shells. Nevertheless, we can extract approximate relations in the sense described in the principles of effective field theory enunciated above. We can presumably describe the physics to any given accuracy in terms of a finite number of parameters. Unfortunately, the number of unknown parameters grows rapidly as we go beyond lowest order in the momentum expansion.

2.2 *Gauge Model Building*

In the early 1970s, after Gerhard 't Hooft's explanation (8) of the renormalizability of spontaneously broken gauge theories, but before the ascendancy of the standard model, many physicists engaged in model building, exploring the huge new space of renormalizable models that 't Hooft opened up to us. This now seems a little naive, but a tremendous effort was expended understanding the range of possibilities for spontaneous symmetry breaking with elementary scalar fields. Much of this effort was devoted to two goals:

1. to determine the precise form of the gauge structure of the partially unified theory of electroweak interactions; and
2. to further unify the electroweak interactions by incorporating their gauge structure into something more comprehensive and simpler.

The first goal, as it turned out, was bootless. Somehow, Glashow, Weinberg, and Salam (9–11) had written down the right electroweak gauge group the first time. The second goal turned out to be more interesting, leading to SU(5) and the idea of grand unified theories (GUTs) (12). This development was very pretty and influential, although we still do not know how close we have actually come to anything related to the real world.

All of the effort in model building had two profound effects on the development of the idea of effective field theory. The obvious consequence was that GUTs made it respectable to think about truly enormous energy scales. I describe the effect of this on the understanding of the renormalization group and related issues in the next three sections. Here I want to mention a less obvious connection. Many workers developed what were actually effective field theory techniques to deal with the complicated scalar potentials that appeared in theories with several different scales.

I describe two related examples of this. The first concerns the so-called hierarchy problem. The issue was whether the large ratios of vacuum expectation values required in GUT models were stable under radiative corrections (13). The answer was yes. A fine-tuning is required to maintain the large ratio, but the required fine-tuning is no worse after radiative corrections than before. The physics of this is an effective field theory. The fine-tuning required is a matching condition onto the low energy effective field theory—the condition (which requires a fine-tuning to implement) is that there must exist a very light Higgs multiplet.

The second example concerns the Glashow-Weinberg condition (14). Glashow and Weinberg discovered that there was a relatively simple way to suppress flavor-changing neutral current effects in SU(2) × U(1) models with more than one scalar doublet. People worried that if there were more than one scalar doublet in the theory, the GIM mechanism (16) would not operate. But Glashow and Weinberg realized that if the doublet that couples to the right-handed quarks with charge 2/3 is distinguished from the doublet that couples to the right-handed quarks with charge $-1/3$ by some symmetry (discrete or softly broken to avoid Goldstone bosons), then the GIM mechanism remains in force even with the extra doublets.

Unfortunately, perhaps because the authors are two of the greatest physicists of the last third of the 20th century, too many readers of the Glashow-Weinberg paper mistook the Glashow-Weinberg condition as a necessary condition on theories with more than one scalar doublet. The common confused argument went something like this. If the Glashow-Weinberg symmetry is not imposed, the neutral components of the two scalar doublets will (in general) have GIM-violating couplings that can produce flavor-changing neutral current effects. Without the symmetry, an unnatural fine-tuning would be required to suppress such couplings. But the neutral components also cannot be very heavy because, like the neutral Higgs field in the standard model with a single fundamental scalar doublet, their masses come from their vacuum expectation values and therefore cannot be much larger than M_W. The flaw in this reasoning is found in the last sentence. If there is no symmetry that distinguishes the two (or more) scalar doublets in an SU(2) $\times$ U(1) model, then we can always choose a basis in the space of the scalar fields so that only one linear combination of the fields gets a nonzero vacuum expectation value (VEV). This one doublet is the only thing that any sensible person could call a "Higgs doublet" in the theory. The other linear combinations of the doublets are simply ordinary scalar fields with positive mass-squared terms, having nothing to do with SU(2) $\times$ U(1) breaking in the model. They are not "Higgs" fields.[2] Nothing prevents them from having masses much larger than M_W. From the effective field theory point of view, it is obvious that flavor-changing neutral current effects in such a theory can be made arbitrarily small by making the masses of the extra doublets very large (15). The effective field theory below the scale of the heavy doublet masses is precisely the one-doublet model, with a GIM mechanism.

You might ask whether this situation requires some special fine tuning. If you think about the scalar doublets in the low energy theory in an arbitrary basis, then it might seem unnatural that the linear combination that gets a VEV should be an approximate mass eigenstate much lighter than all the rest. But from an effective field theory point of view, what you should really do is to work your way down from the higher scales that surely exist in the theory. In this way of thinking,

[2] I like to say that in a theory of this kind, the term "charged Higgs" is an oxymoron.

the situation is reversed. Suppose we look at the matching onto the low energy theory (low as in a TeV, say) from the next higher scale, M, wherever it is. To have any light scalar doublet requires a special fine-tuning. But if you want to have spontaneous symmetry breaking at low scales you are forced to do one such special fine-tuning. If you make one such fine-tuning at the large scale, M, you will typically end up with one light scalar and lots of heavy scalars with mass of order M. Of course, it is the light one that gets the VEV, because it is the only one that survives into the low energy theory in which $SU(2) \times U(1)$ is spontaneously broken. In other words, you actually have to tune to get a small negative mass-squared for the light doublet, while keeping all the other mass-squares positive, to avoid VEVs of order M. The others doublet will get small VEVs of order v^3/M^2, induced by quartic couplings involving three light doublets and one heavy doublet. You can do additional fine-tunings to make more doublets light and end up with a theory with multiple doublets and no symmetry, but then you are doing more ugly fine-tunings than necessary to get the right physics. So the situation in which the doublet with the VEV is nearly a mass eigenstate is, in some sense, the most natural possibility.

2.3 *Renormalization Logs*

A primary reason that effective field theory machinery is nontrivial and useful is the necessity of renormalization in quantum field theory. Indeed, renormalization and effective field theory are closely related ideas. Underlying both is the independence of physics at long distances from the details of physics at short distances. The history of the ideas of regularization and renormalization, the renormalization group, renormalization schemes, and effective field theory are closely tied together.

Some of these connections are obvious. For example, the way that cutoff dependence in a regularized theory gets absorbed into renormalized parameters is closely related to the Appelquist-Carrazone "decoupling theorem" (17) in effective field theory. After all, a cutoff is just a type of new physics at short distances (fake physics, to be sure, because it often violates some of the standard rules and because we do not take it seriously as physical reality). Cutoff dependence gets incorporated into renormalized parameters in exactly the same way in which dependence on a large particle mass gets incorporated into the low energy parameters of the effective field theory describing physics at scales much smaller than the particle mass. Most of this dependence (typically that which goes with a positive power of the cutoff or the heavy-particle mass) is completely absorbed by the low energy parameters and disappears from the low energy physics without a trace, because it involves only the physics at large scales. The important exception is the logarithmic dependence on large scales associated with processes that contribute at all scales. The quantitative upshot of this is that renormalization always involves a dimensional parameter, the renormalization scale, μ, that marks the arbitrary boundary between short-distance physics and long-distance physics in such processes.

Suppose that you integrate out particles with mass m and study physics at $E \ll m$. The situation is complicated because of renormalization. You must choose some renormalization scheme. The scheme must contain additional dimensional parameters, at least the renormalization scale, μ (it could just be a particle mass, but you always need some dimensional scale to renormalize the processes that contribute at all scales). Then there may be terms like

$$\ln \frac{E}{\mu} . \qquad 4.$$

To avoid large logs arising from such terms, you should calculate as much as possible at the renormalization scale $\mu \approx E$. The trouble is that the renormalization scheme can introduce large logarithms, all by itself, of order $\ln m/\mu$. To do accurate calculations, you must find a "good" renormalization scheme, one that avoids these things.

In the full theory, one way to do this is to use a so-called physical renormalization scheme that incorporates the Appelquist-Carrazone theorem (17). Then there will be no dependence on unphysical parameters. A simple example of a physical renormalization scheme is momentum space subtraction in which physical parameters are fixed at particular values of the momenta.

The momentum space subtraction scheme, like any physical renormalization scheme, is mass dependent. This means that anomalous dimensions and β functions must explicitly depend on μ/m. This is what allows the Appelquist-Carrazone theorem to operate. Consider, for example, the Georgi-Quinn-Weinberg (GQW) (18) calculation of coupling constant renormalization in SU(5) (where m_X is the mass of the superheavy gauge bosons). In a renormalization scheme that incorporates the Appelquist-Carrazone theorem, the gauge couplings at scales much larger than m_X will be approximately equal, because the breaking of the SU(5) gauge symmetry has a negligible effect when all the energies in the process are very large compared to m_X. But at energy scales much smaller than m_X, the gauge couplings of the SU(3), SU(2), and U(1) subgroups are very different, each running with a β function determined by low energy physics. We realized that, to leading

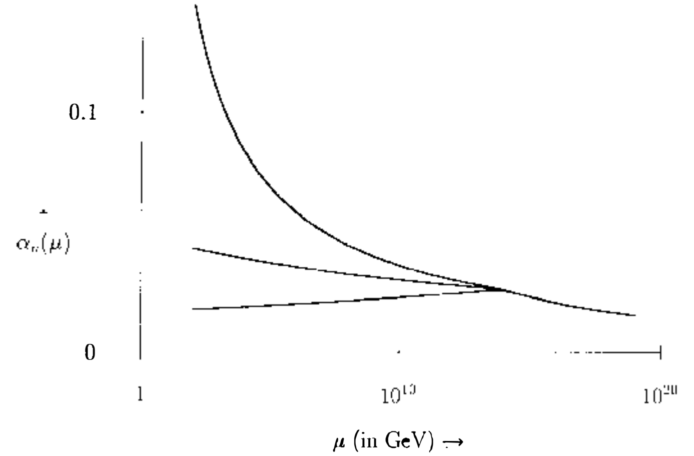

Figure 1 Coupling constant renormalization in SU(5).

order, the couplings at low energies could be related by assuming that they all came together at some large energy scale of order m_X (18). This gives the familiar coupling constant renormalization shown in Figure 1.

The details of the transition from large scales ($\mu \gg m_X$) to small scales ($\mu \ll m_X$) are entirely scheme dependent. However, their effect on the low energy couplings is higher order in α_5 (m_X) because they are important only in a small transition region around $\mu = m_X$.

2.4 *Mass-Independent Schemes*

Some years after the GQW calculation, attempts to improve on it provided a major impetus to the development of continuum effective field theory. Because it seemed necessary to incorporate the Appelquist-Carrazone theorem (19), the first attempts involved two-loop calculations in a momentum space subtraction scheme (20).

Physical renormalization schemes are very intuitive, and of course, "physical." In practice, however, they are difficult to use beyond one-loop in a theory with vcry disparate scales. These difficulties are related to the fact that the renormalization group β functions depend explicitly on the renormalization scale, μ, as well as on the physical parameters.

It is much easier to use a scheme in which all the β functions are independent of μ, and thus depend only on the physical parameters. Such a scheme is called a "mass-independent subtraction scheme."[3]

[3] I believe that this is the most general definition of a mass-independent scheme. It applies even in the nonrenormalizable theories required for an effective theory.

The most familiar and useful examples of mass-independent schemes are dimensional regularization (DR) with minimal subtraction (MS) and its even more useful relative, $\overline{\text{MS}}$. There are many advantages of $\overline{\text{MS}}$. I list a few here and discuss them in more detail below:

1. easier calculations (simpler expressions);
2. automatic "subtraction" rather than "renormalization"; and
3. dimensional analysis.

In spite of these practical advantages, some people still mistrust dimensional regularization because it seems unphysical. In fact, I do not think that there is anything to worry about, so long as DR is used properly. I therefore begin by explaining in what sense DR is a regulator at all.

The physical idea of a regularization scheme is that it is a modification of the physics of the theory at short distances that allows us to calculate the quantum corrections. If we modify the physics only at short distances, we expect that all the effects of the regularization can be absorbed into the parameters of the theory. That is how we chose the parameters in the first place. However, it is not obvious that DR is a modification of the physics at short distances. To see to what extent it is, consider a typical Feynman graph in the unregularized theory in Eucliden space. In one loop (which I discuss for simplicity), all graphs ultimately reduce to sums of objects of the following form

$$I = \int [\mathrm{d}x] \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{1}{(l^2 + A^2)^\alpha}, \qquad 5.$$

where α is some integer.

In DR, these objects get replaced by integrals over $(4 + \delta)$-dimensional momentum space. I am going to think of $\epsilon = -\delta$ as being negative. This doesn't really matter, because everything is defined by analytic continuation anyway, but it makes things easier to talk about. The regularized integrals have the form

$$I_\delta = c(\delta) \int [\mathrm{d}x] \frac{\mathrm{d}^{4+\delta} l}{\mu^\delta (2\pi)^{4+\delta}} \frac{1}{(l_\delta^2 + l^2 + A^2)^\alpha}, \qquad 6.$$

where $c(\delta) \rightarrow 1$ as $\delta \rightarrow 0$, and where I have explicitly separated out the "extra" δ dimensions, so that l^2 is the 4-dimensional length.

In practice, we would do the whole n-dimensional integral at once, using

$$\int \frac{d^n l}{(2\pi)^n} \frac{(l^2)^\beta}{(l^2 - A^2)^\alpha}$$

$$= \frac{i}{(4\pi)^{n/2}} (-1)^{\alpha+\beta} (A^2)^{\beta-\alpha+n/2} \times \frac{\Gamma(\beta + n/2)\,\Gamma(\alpha - \beta - n/2)}{\Gamma(n/2)\Gamma(\alpha)} . \quad 7.$$

However, to see what is happening, I am going to split the integral into one integral over the usual four dimensions and another over the extra δ dimensions. Rewrite the integral as follows:

$$c(\delta) \int [dx] \frac{d^\delta l}{(2\pi\mu)^\delta} \frac{d^4 l}{(2\pi)^4} \frac{1}{(l_\delta^2 + l^2 + A^2)^\alpha} , \quad 8.$$

Now write the integral over δ extra dimensions:

$$I_\delta = \int [dx] \frac{d^4 l}{(2\pi)^4} \frac{1}{(l^2 + A^2)^\alpha} r(\delta) \left(\frac{l^2 + A^2}{4\pi\mu^2}\right)^{\delta/2} , \quad 9.$$

where

$$r(\delta) = c(\delta) \frac{\Gamma(\alpha - \delta/2)}{\Gamma(\alpha)} . \quad 10.$$

The factor $r(\delta) \to 1$ as $\delta \to 0$. The important factor is $\rho^{\delta/2}$, where

$$\rho = \frac{l^2 + A^2}{4\pi\mu^2} . \quad 11.$$

This also approaches unity as $\delta \to 0$, but here convergence depends on l and A:

$$\rho^{\delta/2} = \exp[(\delta \ln \rho)/2], \quad 12.$$

$$\rho^{\delta/2} \approx 1 \qquad \text{for} \qquad |\ln \rho| \ll \frac{1}{\delta} . \quad 13.$$

You can see from Equation 13 that, for very small δ, we have not changed the physics for l (the loop momentum) and A (which involves external momenta and masses) of the order of μ, but that there are significant differences if either l or A is much larger than μ for fixed δ, or if they are both much smaller than μ. The first is exactly what we want. This is just a modification of the physics at short distances. The second is the problem. Dimensional regularization can modify the physics at large distances as well, so that, in general, it is not a sensible regulator.

However, we are all right so long as we avoid infrared divergences. I argue below that effective field theory calculations are very naturally done in $\overline{\text{MS}}$ because infrared divergences are naturally kept at bay. In particular, when a calculation of a matching correction to an effective field theory is properly organized (as outlined below), it is automatically infrared finite. To each order, the matching correction is a difference of two Feynman integrals, one in the high energy theory and one in the low energy theory. The infrared contributions are the same and cancel in the difference.

2.5 *Back to Appelquist-Carrazone*

The decoupling theorem does not work in a mass-independent scheme (21). For example, all the gauge couplings in SU(5) evolve in the same way. This must be true in the SU(5) theory because at scales large compared to the GUT scale there is only one coupling. Since the β functions are independent of μ, this must be true at all scales, even scales much smaller than the GUT scale. The renormalization scale dependence of the Feynman graphs are the same whether there is a gluon or an X particle inside. Thus

$$\frac{12\pi}{\alpha_n(\mu)} \approx (55 - 2f) \ln \frac{\mu^2}{\Lambda^2}. \tag{14}$$

Thus in $\overline{\text{MS}}$, there are large logarithms in low energy calculations in perturbation theory. They must be there to incorporate the physical differences between the different couplings at low energies. But while the renormalization group in a physical renormalization scheme adds up the right infinite set of logarithmic terms to describe this difference, the renormalization group in $\overline{\text{MS}}$ does not. It adds up the wrong infinite set of logarithms.

There are ad hoc ways of dealing with this problem, but the best solution is to use effective field theory. Here the decoupling theorem is not expected to come out—it is put in by hand in matching between the high energy theory and the low energy theory. In each of the two regimes, you can use $\overline{\text{MS}}$, but the theories are different, so you get different renormalizations of the different couplings in the low energy theory. This idea was clearly implied by Ed Witten (22) and stated directly by Steve Weinberg (23). It was used by Lawrence Hall to improve the GQW calculation (24).[4] It was Weinberg's work that really converted me. It is the putting in the physics of Appelquist-Carrazone by hand in switching from one effective theory to another that is the heart of continuum effective field theory.

[4] It was also used by Gilman and Wise in their analysis of the weak Hamiltonian (see next section).

2.6 *Witten, Gilman-Wise, and Heavy Quarks*

At the same time that continuum effective field theory was being elaborated into a useful calculational scheme in the electroweak and GUT arena, something similar was happening in the field of perturbative QCD. Here the motivations were to understand the implications of the heavy or what seemed at the time to be heavy) charmed quark (22, 25) and to have a systematic treatment of the QCD corrections to the electroweak interactions (26).

Witten extended the understanding of the Appelquist-Carrazone theorem by showing in detail how a heavy particle could simply be left out of the low energy theory. Gilman & Wise used this idea and combined it with $\overline{\text{MS}}$ to put the theory of QCD corrections to the electroweak interactions—pioneered by Lee & Gaillard (27) and Altarelli & Maiani (28)—on a simple and firm theoretical footing.

3. "INTEGRATING OUT" VERSUS "MATCHING"

The most interesting and least trivial of the advantages of dimensional regularization with minimal subtraction that I mentioned above is that dimensional analysis works. I want to illustrate what I mean by this in some detail.

In the full theory, the nonlocal result of integrating out heavy particles has unexpected properties with respect to dimensional analysis. In particular, you cannot just Taylor expand in p/m, where p is the momentum and m is the heavy-particle mass. The problem is that light-particle loops can give extra powers of large mass, m. This causes a breakdown of simple dimensional counting (29; see also 17, 22).

For example, imagine that $m_t \gg M_W$. Of course, there are some phenomenological reasons to believe this is not true, but it is fun to think about it nevertheless. Consider the $\bar{b}\not{Z}b$ coupling induced by integrating out a heavy t quark.

If you just integrate out the t, you get a $\bar{b}W\,ZWb$ term proportional to $1/m_t^2$, as shown in Figure 2. You might guess, using dimensional

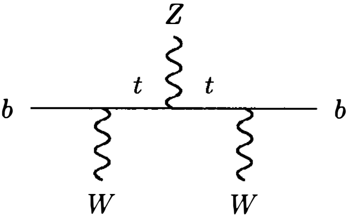

Figure 2 Integrating out the top quark.

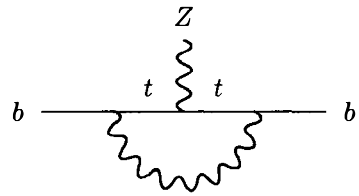

Figure 3 Closing the W line.

analysis that all the effects of this graph are of order $1/m_t^2$. But dimensional analysis does not work, because closing the W loop, as shown in Figure 3, gives terms proportional to m_t^2. However, the m_t^2 term in Figure 3 comes from large momentum in the loop. Therefore, it can be simulated by a local term in the low energy theory. In effective theory, the high momentum physics of Figure 2 is modified so dimensional analysis works, and the m_t^2 part of Figure 3 is put in explicitly by matching.

3.1 *Matching*

The general form of a matching calculation is illustrated in Figure 4. It goes like this. You start at a very large scale, that is with the re-

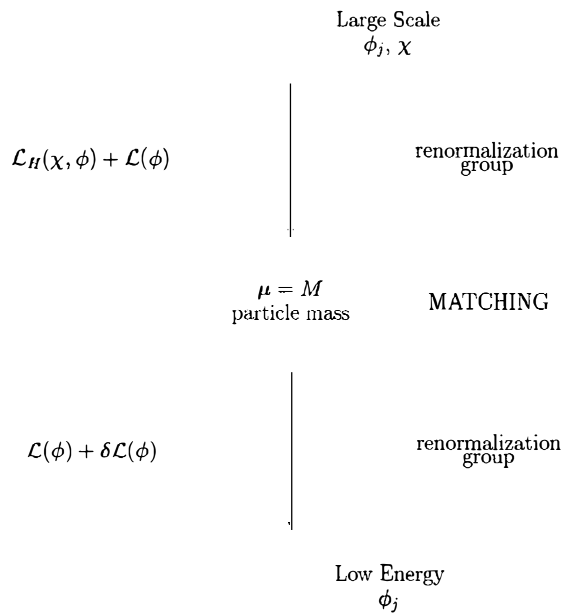

Figure 4 The general form of a matching calculation.

normalization scale, μ, very large. In a strongly interacting theory or a theory with unknown physics at high energy, this starting scale should be sufficiently large that nonrenormalizable interactions produced at higher scales are too small to be relevant. In a renormalizable, weakly interacting theory, you start at a scale above the masses of all the particles, where the effective theory is given simply by the renormalizable theory, with no nonrenormalizable terms.

In this region, the physics is described by a set of fields, χ, describing the heaviest particles, of mass M, and a set of light-particle fields, ϕ, describing all the lighter particles. The Lagrangian has the form

$$\mathcal{L}_{\mathrm{H}}(\chi,\phi) + \mathcal{L}(\phi), \qquad 15.$$

where $\mathcal{L}(\phi)$ contains all the terms that depend only on the light fields, and $\mathcal{L}_{\mathrm{H}}(\chi,\phi)$ contains everything else. You then evolve the theory down to lower scales. As long as no particle masses are encountered, this evolution is described by the renormalization group. However, when μ goes below the mass M of the heavy particles, you should change the effective theory to a new theory without the heavy particles. In the process, the parameters of the theory change, and new, nonrenormalizable interactions may be introduced. Thus the Lagrangian of the effective theory below M has the form

$$\mathcal{L}(\phi) + \delta\mathcal{L}(\phi), \qquad 16.$$

where $\delta\mathcal{L}(\phi)$ is the "matching correction" that contains all the changes.

Both the changes in existing parameters and the coefficients of the new interactions are computed by "matching" the physics just below the boundary in the two theories.[5] It is this process that determines the sizes of the nonrenormalizable terms associated with the heavy particles. Because matching is done for $\mu \approx M$, the rule for the size of the coefficients of the new operators is simple for $\mu \approx M$. At this scale, all the new contributions scale with M to the appropriate power (set by dimensional analysis) up to factors of coupling constants, group theory, or counting factors and loop factors (of $16\pi^2$, etc). Then when the new effective theory is evolved down to smaller μ, the renormalization group introduces additional factors into the coefficients. Thus a heavy-particle mass appears in the parameters of an effective field theory in two ways. There is power dependence on the mass that arises

[5] In a sense, the renormalization group is simply the "matching" of the theory at the scale μ to the theory at the scale $\mu - d\mu$ without changing the particle content.

from matching conditions. There is also logarithmic dependence that arises from the renormalization group.

Generally, the only way we have of calculating matching corrections is in perturbation theory in some small parameter. We analyze it in a loop expansion, which, if there are no factors of $\hbar$ hidden in the parameters of the Lagrangians, is equivalent[6] to an expansion in powers of $\hbar$,

$$\delta\mathcal{L} = \sum_{n=0}^{\infty} \delta\mathcal{L}^n. \qquad 17.$$

We want to choose the matching correction, $\delta\mathcal{L}(\phi)$, so that the physics of the ϕ particles is the same in the two theories at $\mu = M$, just at the boundary.

The strongest form of this equivalence that we might impose is that all one-light-particle irreducible (1LPI) functions with external light particles are the same. By 1LPI, I mean a graph that cannot be disconnected by cutting a single light-particle line. We can describe this in terms of the equality of the light-particle effective actions in the two descriptions. The light-particle effective action in the low energy theory is just the effective action. In the high energy theory, the light-particle effective action is defined as follows. Construct the generating functional for connected Feynman graphs with light-particle external lines, and then Legendre transform it. This gives the generating functional for 1LPI graphs. Then the matching condition is[7]

$$S_{\mathcal{L}_H+\mathcal{L}}(\Phi). \qquad 18.$$

Formally, let us denote the light-particle effective action produced by an arbitrary Lagrangian $\mathcal{L}$ (in the relevant effective theory) by

$$S_{\mathcal{L}}(\Phi) = \sum_{n=0}^{\infty} S^n_{\mathcal{L}}(\Phi), \qquad 19.$$

where n is the number of loops.

[6] There are a number of options here. I regard the Lagrangian in the high energy theory, including the nonrenormalizable terms, as a zero-loop effect, with no internal $\hbar$s. However, one could equally well keep track of where the nonrenormalizable terms come from, in which case the Lagrangian in the high energy theory would also have a nontrivial loop expansion.

[7] A weaker condition could be imposed—that the S matrix for light-particle scattering is the same in the two theories. In this case the light-particle effective actions are not equal, but are related by some redefinition (in general nonlinear) of the fields. I am grateful to Weinberg for emphasizing this. For a good recent discussion, see (30).

At tree level, the light-particle effective action in the low energy theory is just the Lagrangian itself, thus

$$S^0_{\mathscr{L}}(\Phi) = \int \mathscr{L}^0(\Phi) \qquad 20.$$

$$S^0_{\mathscr{L}+\delta\mathscr{L}}(\Phi) = \int \mathscr{L}^0(\Phi) + \int \delta\mathscr{L}^0(\Phi). \qquad 21.$$

In the high energy theory, the light-particle effective action is the sum of all 1LPI tree graphs. Thus it is a sum of the Lagrangian plus all trees with external light particles and only internal heavy-particle lines:

$$S^0_{\mathscr{L}_H+\mathscr{L}}(\Phi) = \int \mathscr{L}^0(\Phi) + \int \begin{Bmatrix}\text{virtual heavy-}\\ \text{particle trees}\end{Bmatrix}(\Phi). \qquad 22.$$

Thus

$$\int \delta\mathscr{L}^0(\Phi) = S^0_{\mathscr{L}_H+\mathscr{L}}(\Phi) - S^0_{\mathscr{L}}(\Phi) = \int \begin{Bmatrix}\text{virtual heavy-}\\ \text{particle trees}\end{Bmatrix}(\Phi). \qquad 23.$$

This is rather trivial, but you can already see that the matching correction is a difference between a calculation in the full theory and one in the low energy effective theory.

The notation here is slightly imprecise. The change in the "Lagrangian," $\int \delta\mathscr{L}^0(\Phi)$, is still nonlocal, in general because it depends on p/M through the virtual heavy-particle propagators. However, it is analytic in p/M everywhere in the region relevant to the low energy theory. Thus it can be expanded in powers with the higher order terms steadily decreasing in importance. It can then be dealt with, to any finite order in the momentum expansion, as a local Lagrangian. In fact, in the low energy theory we always interpret it as if it is expanded to some finite order—that is, as a local Lagrangian.

In the example of integrating out the top quark, the tree diagram of Figure 2 in the full theory gives nonrenormalizable interactions of the form shown in Figure 5. In going to the effective theory and replacing Figure 2 by Figure 5, we have changed the high energy behavior. Dimensional analysis now works. All the effects of Figure 5 are of order $1/m_t^2$.

You might well ask whether we are going to get into trouble because we have modified the high energy behavior in going over to the low energy theory. In particular, effects from the integration of loop momentum that in the full theory are cut off at momentum of order M are now not cut off at all. The way it works is the following. Power law divergences are thrown out in the low energy theory by $\overline{\text{MS}}$, but they

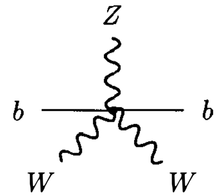

Figure 5 Higher dimension operators from trees.

appear explicitly as terms in the matching corrections, as is discussed below. Logarithmic divergences in the low energy theory are associated with renormalization group running of the parameters. This running starts at $\mu = M$. Thus the renormalization group running in the low energy theory incorporates the $\log M$ terms from the full theory. However, not only are these effects easy to calculate in the low energy theory, because they involve only the infinite renormalization parts of the Feynman graphs, but the renormalization group automatically adds up an infinite sequence of logarithms in the most effective way. Rather than getting us into trouble, this change in the high energy behavior gives us a more efficient way of dealing with the $\log M$ effects.

This situation is even more interesting at one loop. From the general properties of the effective action, we can write

$$S^1_{\mathscr{L}+\delta\mathscr{L}}(\Phi) - S^1_{\mathscr{L}-\delta\mathscr{L}^0}(\Phi) + \int \delta\mathscr{L}^1(\Phi). \qquad 24.$$

This is because the one loop term, $\delta\mathscr{L}^1(\Phi)$, can appear only as a tree diagram. A loop graph with an inclusion of $\delta\mathscr{L}^1(\Phi)$ would give a contribution even higher order in $\hbar$. Thus the matching condition to one loop can be written

$$S^1_{\mathscr{L}_H+\mathscr{L}}(\Phi) - S^1_{\mathscr{L}+\delta\mathscr{L}^0}(\Phi) = \int \delta\mathscr{L}^1(\Phi). \qquad 25.$$

Let us pause to consider what Equation 25 means. Both $S^1_{\mathscr{L}_H+\mathscr{L}}(\Phi)$ and $S^1_{\mathscr{L}}(\Phi + \delta\mathscr{L}^0)$ are nonlocal functions of Φ in which the nonlocality is determined by the long-distance physics. If there are massless particles in the theory, the 1PI functions may not even be analytic in the momenta, p, as $p \to 0$. However, in the difference, all the long-distance physics that gives rise to this nonanalyticity cancels. All the remaining nonlocality is due to the propagation of the heavy particles, and the result can be expanded in powers of p/M with no large coefficients to give the infinite tower of terms in the effective theory. This gives the one-loop contribution to the change in the effective Lagrangian.

Note, in particular, that the matching calculation is infrared finite. An infrared divergence always arises from a loop integration over small momenta. But for small momentum, the full theory and the effective theory give the same physics, by construction. Indeed, if we kept an infinite number of terms in the effective theory, the physics would be exactly the same, and there would be no contribution at all from the loop integral for momenta smaller than M. In practice, we cannot include an infinite number of terms in the low energy theory. To compute the one-loop matching corrections to some order in the momentum expansion, it is only necessary to include enough terms to eliminate all the infrared divergences in the loop integral for the matching correction. If you still have an infrared divergence at one loop to some order in the momentum expansion, it simply means that you have not included enough terms in the tree level matching corrections—you have not gone far enough in the momentum expansion of the tree level matching. This argument works to each order in the loop expansion. The matching that you have already done in the previous order is sufficient to eliminate infrared divergences. Matching calculations are always completely infrared finite.

Note that the same argument implies that the matching corrections are analytic in the parameters of the low energy theory, even light-particle masses. Nonanalyticity in the light masses arises from infrared divergences in loop graphs that emerge when the result is differentiated enough times with respect to the light-particle mass. As we just argued, such divergences cannot arise in a matching calculation. No matter how many times you differentiate with respect to the light-particle masses, if you include enough terms in the momentum expansion at tree level, you will get an infrared-finite matching contribution at one loop. Thus no nonanalyticity in the low energy parameters can ever arise. All nonanalyticity comes from loop effects in the low energy EQFT.

For example, the $\bar{b}\not{t}b$ term is the contribution of the full theory diagram of Figure 3 minus the diagram of Figure 6. All the nonanalyticity associated with the "small" W mass (such as $\ln M_W/m_t$ effects) cancels in the difference. This difference, the one-loop matching contribution, is the missing piece of the full theory calculation, Figure 5, that we threw out when we changed the high energy behavior to make dimensional analysis work.

It is straightforward to go beyond one loop. The only subtlety is that now the lower order changes in $\delta\mathcal{L}$ can also contribute to the long-

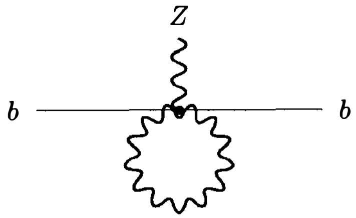

Figure 6 One-loop diagram in the low energy theory.

distance physics and must be subtracted. This can be done using the following:

$$S^n_{\mathcal{L}+\delta\mathcal{L}}(\Phi) = S^n_{\mathcal{L}+\delta\mathcal{L}_{n-1}}(\Phi) + \int \delta\mathcal{L}^n(\Phi), \qquad 26.$$

where

$$\delta\mathcal{L}_k \equiv \sum_{n=0}^{k} \delta\mathcal{L}^n. \qquad 27.$$

The matching condition to n loops can be written

$$\int \delta\mathcal{L}^n(\Phi) = S^n_{\mathcal{L}_H+\mathcal{L}}(\Phi) - S^n_{\mathcal{L}+\delta\mathcal{L}_{n-1}}(\Phi), \qquad 28.$$

or in terms of 1LPI functions,

$$\delta\gamma^{n,j} = \Gamma^{n,j}_{\mathcal{L}_H+\mathcal{L}} - \Gamma^{n,j}_{\mathcal{L}+\delta\mathcal{L}_{n-1}}. \qquad 29.$$

As usual in perturbation theory, we need the lower order results to compute the higher order results.

3.2 *The Operator Product Expansion*

As promised in Section 1, I now discuss what happens if you first integrate out the heavy fields. Formally, we can define the nonlocal effective Lagrangian that results from integrating out the heavy quarks as follows:

$$\int [d\chi] \exp[i\textstyle\int\mathcal{L}_H(\chi,\phi)] = \exp[i\textstyle\int\mathcal{L}_{NL}(\phi)]. \qquad 30.$$

This is formal, because we have not discussed regularization. However, it makes perfect sense in the regularized theory with all counter terms explicitly included. Like $\delta\mathcal{L}(\phi)$, $\mathcal{L}_{NL}(\phi)$ is a nonlocal function of the field ϕ, with the nonlocality determined by the heavy-particle

Compton wavelength, $1/M$. The difference is in the interpretation of the high energy behavior. As discussed in the previous section, $\delta\mathcal{L}(\phi)$ is interpreted as if it is expanded to some finite order in the momentum expansion. Thus $\delta\mathcal{L}(\phi)$ has very bad high energy behavior. But $\mathcal{L}_{\rm NL}(\phi)$ is genuinely nonlocal. It falls off for momenta larger than M.

In terms of $\mathcal{L}_{\rm NL}$, the matching condition is

$$S_{\mathcal{L}+\mathcal{L}_{\rm NL}}(\Phi) = S_{\mathcal{L}+\delta\mathcal{L}}(\Phi). \tag{31}$$

To see what this means in detail, expand $\mathcal{L}_{\rm NL}(\phi)$ in number of loops,

$$\mathcal{L}_{\rm NL} = \sum_{n=0}^{\infty} \mathcal{L}^{n}_{\rm NL}. \tag{32}$$

Now the analog of Equation 22 is

$$\int \delta\mathcal{L}^{0}(\Phi) = \int \mathcal{L}^{0}_{\rm NL}(\Phi). \tag{33}$$

In other words, except for the different interpretation of the high energy behavior, $\mathcal{L}^{0}_{\rm NL}$ is just the tree level matching correction. At low energies, there is no difference. The difference in the high energy behavior, however, is crucial. Because we interpret $\delta\mathcal{L}^{0}$ as truncated after some arbitrary but finite number of terms in the momentum expansion, we can treat $\delta\mathcal{L}^{0}(\phi)$ as a contribution to a local Lagrangian, while $\mathcal{L}^{0}_{\rm NL}$ is nonlocal.

In next order, the matching condition looks like

$$\begin{aligned}\int \delta\mathcal{L}^{1}(\Phi) &= S^{1}_{\mathcal{L}+\mathcal{L}^{0}_{\rm NL}+\mathcal{L}^{1}_{\rm NL}}(\Phi) - S^{1}_{\mathcal{L}+\delta\mathcal{L}^{0}}(\Phi)\\ &= \int \mathcal{L}^{1}_{\rm NL} + S^{1}_{\mathcal{L}+\mathcal{L}^{0}_{\rm NL}}(\Phi) - S^{1}_{\mathcal{L}+\delta\mathcal{L}^{0}}(\Phi).\end{aligned} \tag{34}$$

The breakdown of the operator product expansion analogy can be seen in the last two terms of Equation 34. There can be terms involving more than one power of $\mathcal{L}^{0}_{\rm NL}$ that contribute to $\int \delta\mathcal{L}^{1}(\Phi)$. Only the linear terms are associated with the operator product expansion of $\mathcal{L}^{0}_{\rm NL}$.

4. APPLICATIONS

4.1 *Symmetries at Low Energies*

As a trivial example of the utility of the effective Lagrangian language, we will interpret and answer the following question: What are the symmetries of the strong and electromagnetic interactions?[8] This question requires some interpretation because the strong and electromagnetic interactions do not exist in isolation. Certainly the electromagnetic interactions, and probably the strong interactions as well, are integral parts of larger theories that violate flavor symmetries, parity, charge conjugation, and so on. The language of effective field theory allows us to define the question in a very natural way. Suppose we look at the effective theory at a momentum scale below the W and Z masses and below the masses of whatever particles (besides the longitudinal components of the W and Z) are associated with SU(2) × U(1) breaking. Then only the leptons, quarks, gluons, and the photon appear in the theory. The only possible renormalizable interactions of these fields consistent with the SU(3) × U(1) gauge symmetry are the gauge-invariant QED and QCD interactions with arbitrary mass terms for the leptons and quarks. All other interactions must be due to nonrenormalizable couplings in the effective theory and are suppressed at least by powers of $1/M_W$. Thus it makes sense to define what we mean by the strong and electromagnetic interactions as the QED and QCD interactions in this effective theory. Indeed, this definition is very reasonable. It is equivalent to saying that the strong and electromagnetic interactions are what is left when the weak interactions (and any weaker interactions) are turned off (by taking $M_W \to \infty$).

Now that we have asked the question properly, the answer is rather straightforward. If there are n doublets of quarks and leptons, the classical symmetries of the gauge interactions are an SU(n) × U(1) for each chiral component of each type of quark. These SU(n) symmetries of the gauge interactions do not depend on any assumptions about the flavor structure of the larger theory. They are automatic properties of the effective theory.

Now the most general mass terms consistent with SU(3) × U(1) gauge symmetry are arbitrary mass matrices for the quarks with charge 2/3, quarks with charge −1/3, and the charged leptons plus an arbitrary Majorana mass matrix for the neutrinos.

The Majorana neutrino masses, if they are present, violate lepton number conservation. In fact, such masses have not been seen. If they exist, they must be very small. It is not possible to explain this entirely in the context of the effective SU(3) × U(1) theory, because nothing except lepton number itself forbids such mass terms. Lepton number

[8] The issue of the symmetries of the standard model played an important role in the development of QCD (see, for example, 31).

violation can be imposed, but it does not follow from the structure of theory below M_W. However, if we go up in scale and look at the SU(3) $\times$ SU(2) $\times$ U(1) symmetric theory above M_W and M_Z and if the only fields are the usual fermion fields, the gauge fields, and an SU(2) doublet Higgs fields, then the renormalizable interactions in the theory automatically conserve lepton number.[9] Thus, there cannot be any neutrino mass terms induced by these interactions. If neutrino mass terms exist, they must be suppressed by powers of an even larger mass. Perhaps this is the reason they are so small. It would, of course be a spectacular discovery to measure unambiguously a nonzero neutrino mass. Then we would have some indication of where such a higher scale actually is.[10]

The quark and lepton masses can be diagonalized by appropriate SU(n) transformations on the chiral fields. Thus, flavor quantum numbers are automatically conserved. Strong and electromagnetic interactions, for example, never change a d quark into an s quark. This statement is subtle, however. It does not mean that the flavor-changing SU(2) $\times$ U(1) interactions have no effect on the quark masses. It does mean that any effect that is not suppressed by powers of M_W can only amount to a renormalization of the quark fields or mass matrix, and thus any flavor-changing effect is illusory and can be removed by redefining the fields.

With arbitrary SU(n) transformations on the chiral fields, we can make the quark mass matrix diagonal and require that all the entries have a common phase. We used to think that we could remove this common phase with a chiral U(1) transformation so that the effective QCD theory would automatically conserve P and CP. We now know that because of quantum effects such as instantons this U(1) is not a symmetry of QCD. A chiral transformation changes the θ parameter. Thus, for example, we can perform a chiral U(1) transformation to makc the quark masses completely real. In this basis, the θ parameter will have some value. Call this value $\bar{\theta}$. If $\bar{\theta} = 0$, QCD conserves parity and CP. But if $\bar{\theta} \neq 0$, it does not.

Experimentally, we know that CP is at least a very good approximate

[9] Classically, they conserve lepton number and baryon number separately. The quantum effects of the SU(2) $\times$ U(1) instantons produce very weak interactions that violate lepton number and baryon number but conserve the difference. Conservation of $B - L$ is enough to forbid neutrino masses.

[10] Note, however, that the relation between the higher scale and the neutrino masses depends dramatically on the nature of SU(2) $\times$ U(1) breaking. In a technicolor theory, for example, the neutrino masses are typically much smaller because they arise from higher dimension operators in which fermion bilinears replace the Higgs doublets.

symmetry of QCD. The best evidence for this comes from the very strong bond on the neutron electric dipole moment, less than 6×10^{-25} e-cm and going down. This implies that $\bar{\theta}$ is less than about 10^{-9}, a disturbingly small number. This situation is sometimes referred to as the strong CP puzzle. The puzzle is: why is $\bar{\theta}$ so small if the underlying theory violates CP?

4.2 *Composite Higgs Bosons*

As in the previous section, we must first decide on a point of nomenclature. What is a composite Higgs boson? Well, what is a Higgs boson? While the term is sometimes loosely applied to any random scalar meson, as discussed in Section 2.2, this is not very appropriate. I begin by discussing the issue in the simplest version of standard model, with a single fundamental doublet of scalar fields and comparing it with the simplest technicolor model. Part of the confusion stems from the fact that Peter Higgs's name is associated with two phenomena that do not always go together.

The "Higgs mechanism" is the process in which a Goldstone boson produced by the spontaneous breaking of a continuous symmetry and the massless gauge particle associated with the gauged unbroken symmetry fuse into a massive gauge boson. The Higgs mechanism occurs any time a gauged continuous symmetry is spontaneously broken. The Higgs mechanism exists and is relevant to the world. We have seen the Higgs mechanism in action in the massive W and Z. But the very generality of the Higgs mechanism implies that its existence does not tell us anything about the mechanism of spontaneous symmetry breaking. We know that SU(2) $\times$ U(1) symmetry has been spontaneously broken, but we do not know what the physics of that breaking is. We do not know what the Goldstone bosons eaten by the Higgs mechanism are made of.

The "Higgs boson," on the other hand, is one possible answer to the question of what breaks the SU(2) $\times$ U(1) symmetry. If the Goldstone bosons are part of a complete SU(2) doublet that develops a vacuum expectation value, then a neutral scalar survives after the Higgs mechanism as an observable state in the low energy theory. The Higgs boson has several precisely defined properties because it is part of a doublet with the Goldstone bosons. For example, its coupling to longitudinal gauge boson pairs is fixed. In fact, in a theory with more than one doublet, the Higgs boson may not be a mass eigenstate. The scalar that couples to $Z^{\mu}Z_{\mu}$ may be a linear combination of several mass eigenstates. Nevertheless, it is this linear combination that is really the Higgs boson. Its coupling to $Z^{\mu}Z_{\mu}$ is fixed. Each of the neutral mass eigenstates can be regarded as a well-defined fraction of a Higgs boson, with the fraction being the absolute square of the ratio of the coupling of the particle to $Z^{\mu}Z_{\mu}$ to the coupling of the Higgs boson in the one-doublet model.

To talk about a Higgs boson at all, you must be able to identify the transformation properties of the Goldstone bosons under the full SU(2) $\times$ U(1) symmetry. This is not always possible. For example, in a technicolor model, it need not make any sense. Very likely, the bound states of the technifermions cannot be classified according to their transformation properties under SU(2) $\times$ U(1) any more than the light hadrons in QCD can be classified under their transformation properties under chiral SU(2) $\times$ SU(2).[11] In such a technicolor model, there is no Higgs boson. The Higgs mechanism operates as usual. The Goldstone bosons are bound states of technifermion and antitechnifermion. But it makes no sense to classify the Goldstone bosons as part of a multiplet of the unbroken symmetry. There is no scalar state in the theory with anything like the properties of the Higgs boson.

There may be an intermediate between the fundamental scalar doublet that leads to a fundamental Higgs boson and technicolor that gives no Higgs boson at all. A class of theories exists in which there is a Higgs boson with much the same properties as the fundamental Higgs boson, but it is a composite state. To construct such a theory, you must have the SU(2) $\times$ U(1) symmetry survive below the scale of the new strong interactions that bind the Goldstone bosons. Models of this kind were first constructed many years ago by David Kaplan and me (33). We called the new strong interactions "ultracolor," and we showed that we could construct such models with ultracolor that was just an ordinary strong gauge interaction like QCD.

The effective field theory idea enters in two ways into these models. First, the Higgs doublet is a pseudo-Goldstone boson associated with a spontaneously broken chiral symmetry of the ultracolor model. In the simplest model, the ultracolor model is analogous to QCD with three light quarks, SU(2) $\times$ U(1) is analogous to the isospin $\times$ hypercharge subgroup of Gell-Mann's SU(3), and the Higgs doublet is the analog of the K meson doublet. Because the Higgs doublet is a pseudo-Goldstone boson, we could use a chiral Lagrangian approach to study its properties. In particular, we could show that with suitable explicit chiral symmetry-breaking interactions, we could produce a light doublet with a $\lambda\phi^4$ interaction.

[11] This is obvious if the chiral symmetry-breaking phase transition is first order. The situation is a little cloudy if the chiral symmetry-breaking transition is second order (see 32).

The second way in which the effective field theory idea enters in these models is to show that, when the chiral symmetry-breaking dynamics is fine-tuned to produce a scalar doublet that is very light compared to the compositeness scale, the effective field theory at low energies looks approximately like the standard model. For example, even if the ultracolor dynamics does not have a custodial SU(2) symmetry, (34) the effective low energy theory still does, at least approximately, because it is just the standard model with small nonrenormalizable corrections.

More recently, the idea of composite Higgs bosons has been rediscovered in the context of models in which the Higgs boson is a bound state of t and $\bar{t}$ (35). These models are superficially attractive because the Higgs boson is a bound state of particles that are expected to exist for other reasons, rather than of completely new states. Unfortunately, the nature of the ultracolor dynamics is not as well understood here as it was in the earlier model (33; cf 36). However, in all these models, as in the original composite Higgs models, the low energy dynamics is simply the dynamics of the standard model with the effects of the high energy dynamics that binds the Higgs doublet put in as a boundary condition (37).

4.3 *Above the* Z

One of the most important applications of effective field theory technology today is to the issue SU(2) × U(1) breaking. The general question here is the following: What does the physics we see at scales up to and just above the masses of the W and Z tell us about higher scales that we cannot see directly? Even before the discovery of the W and Z, the observed properties of the weak interactions of quarks and leptons convinced almost all particle physicists that they must exist, as the massive gauge bosons of spontaneously broken SU(2) × U(1). Amazingly, this history seems to have been forgotten by some. One still occasionally sees papers in which the properties of the W and Z are discussed without proper regard to the constraints of SU(2) × U(1) symmetry. Thus it may be useful to recount the important issues.

The story of the weak interactions was based on an effective field theory from the beginning. The nonrenormalizable four-fermion interactions that accurately describe the charged current weak interactions at low energies contain the dimensional parameter

$$v = \sqrt{\frac{1}{\sqrt{2}G_F}} \approx 250 \text{ GeV}. \qquad 35.$$

Thus it was clear that four-fermion theory of the weak decays was an effective theory that would have to be replaced by new physics at high energies.

The current-current form of the charged current weak interactions was an early clue that the new physics would have something to do with gauge invariance, and the connection with the electromagnetic interactions suggested an SU(2) $\times$ U(1) with massive W and Z gauge bosons (9). But the theory of the massive W and Z coupled to the nonconserved weak currents is still not renormalizable. It still contains the dimensional parameter, v, now hidden in the couplings of the longitudinal spin component of the W and Z. In a modern effective field theory language, we would say that the appearance of v in the effective theory is a signal that there must be new physics at some higher scale, $\Lambda \lesssim 4\pi v$ (4, 38). The same result was obtained in other ways in the 1970s (39).

Weinberg (23), Salam (11), and 't Hooft (8) showed us how to incorporate the Higgs mechanism to make sense of the electroweak interactions at scales above $4\pi v$. In the 1970s, experimenters collected a tremendous body of experimental evidence for the simplest embedding of quarks and leptons into SU(2) $\times$ U(1). While the picture will not be complete until the experimental discovery of the t quark and the experimental clarification of the mechanism of *CP* violation, we can be reasonably confident that physics above the Z is a spontaneously broken SU(2) $\times$ U(1) gauge theory with all left-handed quarks and leptons in doublets and all right-handed quarks and leptons in singlets, at least up to scales of the order of a few TeV.

The joke in this description of physics above the Z is that the existence of spontaneously broken SU(2) $\times$ U(1) does not tell us what actually breaks the SU(2) $\times$ U(1) symmetry. So long as the physics of SU(2) $\times$ U(1) breaking preserves a custodial SU(2) symmetry, we get the same zeroth-order picture of the electroweak interactions (34, 40).

Why do I say that this picture works up to scales of order $4\pi v$, a few TeV, even though we have only observed it directly up to scales of order 100 GeV? Because if there were new physics at some scale, M, much less than $4\pi v$, we would expect it to modify the predictions of the standard model at energies of the order of M_W by terms of order

$$\frac{M_W^2}{M^2} \gg \frac{e^2}{16\pi^2 \sin^2\theta}. \qquad 36.$$

In other words, if $M \ll 4\pi v$, the effects of the new physics will be much greater than the typical size of radiative corrections in the standard model. Because we do not see any dramatic departures from standard model physics at this level, we can reasonably assume that there is no important new physics besides SU(2) × U(1) breaking below a few TeV. If there were new physics in this region with significant couplings to the fields of the standard model, there would be no reason for the large, complicated, and beautiful picture of the standard model to hang together as it does.

Note that it is only because the gauge coupling is small that we can give a detailed description of the properties of the massive vector bosons without a similarly detailed account of the physics of SU(2) × U(1) breaking. It is the fact that $e^2/(16\pi^2 \sin^2 \theta)$ is much less than one that implies a large gap between the mass of the W and Z and the scale by which the physics of SU(2) × U(1) breaking must appear. Because of this large gap, dependence on the details of SU(2) × U(1) breaking is suppressed. Of course, this is really just the statement that we know in detail what is going on only in perturbation theory. For large coupling, the typical radiative corrections are large and it is much less meaningful to say that a vector boson is associated with a spontaneously broken gauge symmetry.

Nevertheless, it should be clear that the statement that the W and Z are the massive gauge bosons associated with spontaneous SU(2) × U(1) breaking is not at all trivial. It certainly does not follow from general principles of effective field theory. It is true that one can formally write any vector boson field as the massive gauge boson associated with a spontaneously broken symmetry, nonlinearly realized (see 42, 43). However, to make a physically meaningful statement, rather than something that is empty mathematics, you need much more. You must be able to identify the gauge coupling, and it must be small, so that the gap between the mass of the vector particle and the scale of symmetry breaking is large. Furthermore, the structure of spontaneously broken gauge symmetry must persist all the way to the scale of spontaneous symmetry breaking. These requirements are apparently satisfied in the electroweak interactions.

Of course, it is logically possible that there is some other new physics just above the Z besides the physics of electroweak symmetry breaking that has been carefully fine-tuned to avoid detection in all experiments done to date. But it is probably not necessary to try to imagine all the ways in which this might be done unless some evidence for new physics actually emerges from experiment. Instead, we should concentrate on pushing on beyond the scale $4\pi v$.

4.4 *Resumming Large Radiative Corrections*

The application of effective theory technology to the issue of SU(2) × U(1) breaking has become a huge industry in the last few years (44). Here I give only one simple illustration of the power of effective field theory in dealing with the standard model at a scale $\mu \approx M_Z$. I focus on the issue of "resummation" of large radiative corrections (53). I must begin by making more precise the discussion of the physics of electroweak symmetry breaking in Section 4.3. I assume that whatever the new physics may be, it shows up only at energies larger than the W and Z masses, so that it is not directly included in the effective field theory at $\mu \approx M_Z$. Thus the effective theory will contain only the quarks and leptons and the massive gauge bosons. I make two assumptions about the physics of the SU(2) × U(1) breaking scale that are crucial in determining the structure of the effective theory.

1. The light quarks and leptons do not participate directly in the physics of SU(2) × U(1) breaking, except through the interactions that give rise to their masses. In particular, I assume that there are no four-fermion operators involving only the light quarks and leptons coming from physics at this scale, except those proportional to powers of the masses. The success of the standard model so far suggests strongly that any four-fermion operators not proportional to the masses arise, if at all, from a considerably larger scale.
2. The SU(2) × U(1) breaking interactions have a custodial $SU(2)_C$ interaction in the limit in which the SU(2) × U(1) gauge couplings are turned off. The U(1) gauge boson, B^μ, transforms like the neutral component of an $SU(2)_C$ triplet, and thus breaks the $SU(2)_C$ in a specific way for nonzero gauge coupling. Furthermore, I assume that there is no larger global symmetry of these interactions than the $SU(2)_L \times SU(2)_C$, so that there are no pseudo-Goldstone bosons. Thus I assume that the symmetry structure of the low energy theory is that of

$$\frac{SU(2)_L \times SU(2)_C}{SU(2)_{L+C}} \,. \qquad 37.$$

This assumption can be easily relaxed to include the effects of additional pseudo-Goldstone bosons and more interesting embeddings of SU(2) × U(1) into a technicolor global symmetry group (e.g. 45–47). Indeed, one of the advantages of the effective field theory formalism is that these generalizations can be organized in a systematic way. However, I do not discuss them here. In the interests of clarity, I stick to the minimal possible assumptions.

The above assumptions are satisfied in the simplest version of the standard model, with a single light doublet of spinless bosons (48) or the simplest sort of technicolor theory with the left-handed technifermions transforming as a single doublet under $SU(2)_L$, or in some other theory with the same symmetry structure.

Sometimes, when it is important to keep the symmetries of the theory manifest, or when estimating the size of higher dimension terms in the effective Lagrangian, I keep the Goldstone boson fields explicitly in a nonlinear realization of Equation 37, described by the unitary field,[12]

$$U \equiv \frac{\sqrt{2}}{v}\begin{pmatrix}\tilde{\Phi}^\dagger \\ \Phi^\dagger\end{pmatrix}, \qquad UU^\dagger = U^\dagger U = 1, \tag{38.}$$

with $\tilde{\Phi} = i\Phi^*\tau_2$ and $v \approx 250$ GeV is the $SU(2) \times U(1)$ breaking vacuum expectation value. The nonlinear field U transforms under $SU(2)_L \times SU(2)_C$ as

$$U \rightarrow C\, U\, L^\dagger, \tag{39.}$$

where C and L are the unitary 2×2 matrices defining the $SU(2)_C$ and $SU(2)_L$ transformations. The nonlinear field Φ then transforms under $SU(2) \times U(1)$ like a conventional Higgs doublet, but it satisfies

$$\Phi^\dagger\Phi = \frac{v^2}{2}\,. \tag{40.}$$

In this notation, the leading terms in the effective Lagrangian look exactly like the renormalizable couplings in the standard model with a fundamental Higgs doublet, except that the scalar potential is not there.

However, we will also need to identify the physical states in the theory, and for this purpose it is more convenient to work in unitary gauge, which corresponds to $U = I$. It is easy to go back and forth from one description of the effective theory to the other (see below).

In the analysis of the effective theory, we concentrate on the dependence on m_t, the t quark mass. The fermions are singlets under the custodial $SU(2)_C$ symmetry. The quark and lepton mass terms, including m_t, break the separate flavor symmetries of the quark and lepton kinetic energy term. That is their function. However, they break the flavor symmetries in a very specific way. They also break the custodial $SU(2)_C$ symmetry.

If $m_t \approx M_Z$, then a reasonable strategy is that of Jenkins & Manohar

[12] I use here the notation of Gasser & Leutwyler (6), who have done the most careful and systematic study of the application of chiral Lagrangian techniques to QCD.

(49), to eliminate the t and the W and Z from the theory at the same scale. However, if $m_t \gg M_Z$, this scheme misses important physics. Likewise, the Jenkins-Manohar scheme is inappropriate for dealing with custodial SU(2) violation from very heavy multiplets. Thus we consider the effect of explicit custodial SU(2) violation in the effective theory below m_t as an example of the effect of another mass scale on the effective field theory formalism. Now there are additional contributions to the parameters that appear to leading order in the derivative expansion. The most important of these is the contribution to the W_3 mass that arises when the t quark is integrated out of the effective theory. We can incorporate this effect by including spurion fields, $P_\pm$, that project onto the $T_3 = \pm 1/2$ components of the $SU(2)_C$ index of the U field. The largest effect is on the custodial SU(2)-violating W_3 mass term, from the operator

$$\frac{1}{2}\,\delta M_3^2\,\frac{1}{16\pi^2}\,[\mathrm{tr}\,P_+\;U\,D^\mu\,U^\dagger]^2, \qquad 41.$$

where

$$P_+ = \frac{1+\tau_3}{2}\,. \qquad 42.$$

We expect δM_3^2 from the integration out of the t to be of order m_t^2 [or in general, of the order of the mass-squared of the particle that is integrated out to produce the custodial SU(2) violation].

The contribution to the ρ parameter from the t in the conventional effective field theory language (see 50 for an effective field theory discussion based on the original calculation of Veltman) is

$$\delta\rho = \frac{3\alpha m_t^2}{16\pi s^2 M_W^2}\,, \qquad 43.$$

(with $s^2 = \sin^2\theta_W$, where θ_W is the weak mixing angle), which implies

$$\delta M_3^2 = -m_t^2\,\frac{3}{4}\,. \qquad 44.$$

Note that although this is a radiative correction in the sense that it comes from a one-loop diagram (the t quark loop), there is no sense in which the contribution is order α, for a heavy t.

When doing precise calculations of radiative corrections, it helps to identify the largest potential contributions to whatever you are calculating and deal with them nonperturbatively, if possible. The effective field theory allows us to do this with the ρ parameter. The important point is that not only can $\delta\rho$ be large compared to most radiative corrections, but it enters into the effective theory in a very special way—as a term in the gauge boson mass matrix.

In the limit of keeping only $\delta\rho$, the relevant oblique[13] terms in the effective field theory renormalized at the Z scale can be written (in unitary gauge) as

$$-\frac{1}{4}W_a^{\mu\nu}W_{a\mu\nu}-\frac{1}{4}B^{\mu\nu}B_{\mu\nu}+\frac{(1+\delta\rho)}{2}\overline{M}_Z^2\,[\bar{c}^2\,W_j^{\mu}W_{j\mu}+(\bar{c}W_3^{\mu}-\bar{s}B^{\mu})(\bar{c}W_{3\mu}-\bar{s}B_{\mu})]-\frac{1}{2}\delta\rho\,\overline{M}_Z^2\,(\bar{c}W_3^{\mu}-\bar{s}B^{\mu})(\bar{c}W_{3\mu}-\bar{s}B_{\mu}), \qquad 45.$$

where the sum over a runs from 1 to 3 and the sum over j from 1 to 2. The parameter $\overline{M}_Z$ is the Z mass parameter renormalized at $\mu \approx M_Z$. The parameters $\bar{s}$ and $\bar{c}$ ($\bar{s}^2+\bar{c}^2=1$) are defined in terms of the running SU(2) and U(1) gauge couplings, g_2 and g_1, renormalized in $\overline{\mathrm{MS}}$ at $\overline{M}_Z$: $g_2=\bar{e}/\bar{s}$, $g_1=\bar{e}/\bar{c}$. From the form of Equation 45, we see that up to small, finite corrections

$$M_Z^2=\overline{M}_Z^2,\qquad M_W^2=\bar{c}^2\,(1+\delta\rho)\overline{M}_Z^2,\qquad \bar{e}(M_Z)=e(M_Z). \qquad 46.$$

While this is very simple, it actually contains an important nonperturbative insight. To see this, imagine what happens to this effective theory when the t quark gets very heavy (of course, the conventional wisdom is that this leads to phenomenological problems, but ignore that for now). Now, in the Higgs model of SU(2) $\times$ U(1) breaking, the Yukawa coupling of the t quark, f_t, is large, and we cannot ignore the effect of Higgs exchange in the calculation of radiative corrections. Nevertheless, the only place we get a contribution to any oblique correction that is proportional to $\alpha_t \equiv f_t^2/4\pi^2$, with no extra factor of α, is from the coefficient $\delta\rho$. This follows from the argument of Cohen et al (50). The main result of that work was the effective field theory explanation of the large correction to the ρ parameter in electroweak interactions from a heavy t quark. It follows from dimensional analysis that the dimension-2 operator (Equation 41) can get a contribution proportional to m_t^2 when the t quark is integrated out of the effective theory. This is the leading contribution to custodial SU(2)-violating oblique effects because it is a contribution of order

$$\alpha_t\cdot F(\alpha_t), \qquad 47.$$

[13] Oblique in this context means contributing only to the gauge boson propagators [45, 51; D. Schildknecht directed my attention to his paper (52), which contains related material].

where $F(\alpha_t)$ describes the effects of the multiple scalar exchange. All other contributions are nonleading, of order

$$\frac{\alpha}{s^2\pi}\cdot G(\alpha_t) \tag{48.}$$

or smaller. If suitably normalized, the function F describes a ratio, the full nonperturbative contribution to ρ over the leading one-loop contribution. This ratio is expected to be of order 1, but it could just as well be 2 or 0. This is the sense in which we incorporate nonperturbative information by working directly with the parameter $\delta\rho$ in the effective low energy theory. Although we cannot reliably calculate $F(\alpha_t)$ or $G(\alpha_t)$ in perturbation theory for large α_t, we can collect all the leading contributions into the parameter $\delta\rho$ and guess the size of other contributions. This is the power of the effective field theory.

To get something out of this, we need to set the scale by comparing with the strength of the low energy weak interactions. When the W is integrated out of the theory at the W scale, it produces a set of four-fermion operators with a coefficient proportional to

$$\frac{\bar{e}^2}{8\bar{s}^2(1+\delta\rho)M_W^2} \approx \frac{\bar{e}^2}{8\bar{s}^2\bar{c}^2(1+\delta\rho)M_Z^2}, \tag{49.}$$

where $\bar{e}^2$ is the electromagnetic coupling renormalized at the Z scale.[14] The contribution to μ decay is further modified by finite radiative corrections at low energies, so we have a relation of the form

$$\frac{G_F}{\sqrt{2}} \approx \frac{\bar{e}^2}{8\bar{s}^2\bar{c}^2(1+\delta\rho)M_Z^2}. \tag{50.}$$

Thus

$$\bar{s}^2\bar{c}^2 \simeq \frac{\pi\alpha_*}{\sqrt{2}G_F(1+\delta\rho)M_Z^2}, \tag{51.}$$

where $\alpha_* = \alpha(M_Z)$, or

$$\bar{c}^2 = \frac{1}{2}\left[1+\sqrt{1-\frac{4A}{(1+\delta\rho)M_Z^2}}\right], \tag{52.}$$

where

$$A = \frac{\pi\alpha_*}{\sqrt{2}G_F}. \tag{53.}$$

[14] In principle, we could distinguish between the W and Z scales in all of this, but because $\bar{c}^2$ is close to 1, this doesn't matter very much, and we ignore the distinction.

Combining Equations 52 and 46, we find the "nonperturbative" relation

$$M_W^2 = \frac{(1+\delta\rho)M_Z^2}{2}\left[1+\sqrt{1-\frac{4A}{(1+\delta\rho)M_Z^2}}\right]. \qquad 54.$$

Except for small corrections, this is the relation found by Lynn, Peskin, and Stuart (51) and reviewed by Consoli (53). We now see that this "conjecture" was actually proven implicitly by Cohen et al (50). It follows rather trivially from the fact that $\delta\rho$ is the largest correction term in the effective field theory for large m_t.

Note also that the proof does not involve any complicated discussions of schemes. The usual apparatus of renormalization has been taken apart into two steps. First the infinities are removed by the automatic procedure of $\overline{\text{MS}}$ renormalization, making no reference to physical parameters, but incorporating all the right dependence on the scales in the process, because we are in the right effective field theory. Then the $\overline{\text{MS}}$ parameters are related to physical parameters.

For large m_t, we must also worry about the nonoblique corrections. Let us find all the terms in the effective Lagrangian that can be enhanced by powers of m_t in the sense that the leading, largest contribution to the term is proportional to a positive power of m_t. We have already seen one example of such a term in Equation 41, the t quark contribution to ρ. We now give the complete list of such terms in the effective theory. The important point is that from the point of view of the effective low energy Lagrangian, the t mass term is a coupling of the Goldstone bosons to the t_R and the left-handed doublet containing the t_L. For each Goldstone boson field we expect a factor of the Yukawa coupling, of order m_t/v. Then the overall scale of the term is set by dimensional analysis using the mass scale m_t, and we must include appropriate loop factors. But we must not forget the symmetry structure. Each appearance of the Yukawa coupling is accompanied by the spurion field that describes the breaking of the flavor symmetry and custodial $SU(2)_C$ symmetry by the mass term. Several simple but important remarks can be made.

When the t quark is integrated out of the theory, the fact that it was part of an SU(2) × U(1) doublet with the b'_L quark (where $b' = \sum_{q=b,s,d} V_{tq} q$ is the appropriate CKM linear combination) is reflected in the structure of the effective theory. In a gauge in which the SU(2) symmetry is manifest, and in the basis in which we label the left-handed quark doublets by their charge-2/3 quark components, the "doublet", ψ_{tL}, containing the b'_L must transform nonlinearly. The field satisfies the nonlinear constraint

$$P_+ \, U \, \psi_{tL} = 0 \qquad 55.$$

or

$$P_- U \psi_{tL} = U \psi_{tL} \qquad 56.$$

that, in unitary gauge, just reduces to the statement that the t quark is absent. Any coupling to quarks that is enhanced by factors of m_t and not suppressed by other quark or lepton masses must involve only the left-handed doublet ψ_t, that contained the t quark.

In this basis, the counting of powers is relatively simple. The combinations $U \psi_{tL}$ or U alone appear with a factor of m_t, because the mass term had the form $m_t \overline{\psi_{tR}} U \psi_{tL}$. The field ψ_{tL} by itself appears with a factor m_t/v, because U depends on the combination Φ/v so we get an extra factor of $1/v$ when we integrate out the Goldstone boson field, or equivalently, because the Yukawa coupling is of order m_t/v. We get the maximum enhancement from a contribution to an operator with the lowest dimension, because extra dimensional factors in the numerator are compensated by factors of m_t in the denominator. We can organize the terms by their number of ψ_{tL}'s.

Four ψ_{tL}'s: The lowest dimension term possible is the GIM-violating four-fermion interaction

$$\frac{m_t^4}{v^4} \frac{1}{16\pi^2} \frac{1}{m_t^2} \overline{\psi_{tL}} \gamma^\mu \psi_{tL} \overline{\psi_{tL}} \gamma_\mu \psi_{tL}, \qquad 57.$$

where I have shown all the relevant factors in the coefficient explicitly—one m_t/v for each of the four ψ's, a $1/16\pi^2$ loop factor, and a $1/m_t^2$ required by dimension counting. This is the usual GIM-violating term from the box diagram. Evidently, it just manages an enhancement of m_t^2. All higher dimension operators involving four ψ_{tL}'s are not enhanced because they have additional $1/m_t^2$ factors from dimension counting. Note that the U plays no role here, because the matrix U is unitary. It is only relevant when differentiated (this is another way of saying that it describes a Goldstone boson field with only derivative interactions), but derivatives increase the dimension of the operator and kill the m_t^2 enhancement. Of course, extra factors of $m_t U$ do not help because each such factor increases the dimension and the extra m_t is compensated by a $1/m_t$ from dimension counting. Also, there are no enhanced operators with more than four ψ_{tL}'s, because the dimension of the operators is too high.

Two ψ_{tL}'s: Since we have only two factors of m_t to play with, no factors of m_t can be tolerated in the denominator, and thus only operators of dimension 4 or lower can be enhanced. But there are no operators with dimension less than 4 and two ψ_{tL}'s. There are two dimension-4 operators involving two ψ_{tL}'s. One is the kinetic energy term. But this already has an order-1 contribution, so the m_t^2 dependence is not leading. However, the term

$$\frac{m_t^2}{v^2}\frac{1}{16\pi^2}\,\mathrm{tr}\,[P_+\,(D^\mu U)U^\dagger]\,\overline{\psi}_{tL}\,\gamma_\mu\,P_-\,\psi_{tL} \qquad 58.$$

is enhanced. In unitary gauge, this is a custodial SU(2) and GIM-violating coupling of the Z to $b\overline{b}$. It is clear that the contribution of this term to the hadronic width of the Z can be used to provide additional bounds on large m_t, and that it will require two miracles, rather than only one, to live with a very large m_t in the nonperturbative regime.

Zero ψ_{tL}'s: Here we must have at least two $m_t D^\mu U$'s, so to get enhancement, we must stop at dimension 4 (as before, extra $m_t U$'s do not help). There are two terms, the $SU(2)_C$ invariant kinetic energy term, which is a nonleading contribution, and the term of Equation 41 that contributes to ρ.

These three operators, containing four, two, or zero ψ_{tL}'s, contain all the leading m_t^2 dependence in the effective theory below m_t. The nonleading contributions will also depend on m_t, because their coefficients will, in general, be some power series in the t quark Yukawa coupling, m_t/v, like Equation 48. But these three are the only operators that begin with a term proportional to m_t^2.

4.5 *Nonperturbative Matching Corrections*

It seems likely to me that the formal picture of matching presented in Section 3.1 in the loop expansion can be extended to give information about nonperturbative contributions to matching. The basic idea is that for nonperturbative as well as perturbative matching, the contribution will be a difference between a full theory calculation and an effective field theory from which all low momentum contributions cancel. Thus, like the perturbative contributions, the nonperturbative will appear local at distances below the matching scale, and can therefore be incorporated into local matching corrections.

To see how this works in more detail, consider the functional integral representations for the vacuum-to-vacuum amplitudes in the two theories. Nonperturbative effects are associated with nontrivial field configurations—with fields that are far from zero in the functional integral. Because only large momentum physics will contribute to the matching corrections, the relevant nontrivial field configurations in the matching corrections will be small almost everywhere, with the nontriviality confined to small regions of size $1/M$ or smaller. Thus we can use the dilute gas approximation (54). In this approximation, the leading contributions to the light-particle effective action come from single, isolated small field configurations integrated over relevant zero-mode parameters.

In the language of Section 3.1, we can distinguish two possibilities for nontrivial field configurations.

1. If the heavy fields, χ, are nontrivial, then the size of the nontrivial field configurations will be set by M. Such field configurations contribute directly to the matching corrections. Obviously, there are no corresponding configurations in the low energy theory to subtract, because the heavy fields do not exist in the low energy theory. In this case, it is obvious that only small configurations contribute to matching, because only small configurations contribute at all.
2. If the nontrivial field configuration involves only the light fields, ϕ, then there are corresponding large nonperturbative effects in the high energy and low energy theories. This is the situation, for example, in QCD when a heavy quark is integrated out of the theory. The nontrivial gluon field configurations contribute in both the high energy and the low energy effective theories. This requires further discussion.

For large light field configurations of size greater than $1/M$, the contribution in the high energy and low energy theory are the same, so there is no contribution to the matching correction. However, the small field configurations have completely different properties in the high and low energy theories.

In the low energy theory, the functional integral has the form

$$\int [d\phi] \exp[i\textstyle\int(\mathcal{L} + \delta\mathcal{L})]. \qquad 59.$$

The nonrenormalizable terms in $\delta\mathcal{L}$ have bad high energy behavior for momenta greater than M. The action is large and rapidly varying for light field configurations of size smaller than $1/M$. Thus in the low energy theory, there are no important nonperturbative contributions from field configurations of size smaller than $1/M$.

In the high energy theory, the functional integral has the form

$$\int [d\phi] \exp[i\textstyle\int(\mathcal{L} + \mathcal{L}_{NL})]. \qquad 60.$$

The physics of the heavy particles in $\mathcal{L}_{NL}$ makes a large contribution only for momenta of order M. The contribution of large field configurations is the same as in the low energy theory and, as expected, cancels from the matching correction. The leading nonperturbative contribution to the matching correction thus comes only from configurations of size less than $1/M$.

For example, in QCD at a heavy quark mass, m, the above argument implies that the nonperturbative matching correction is the conventional instanton contribution from instantons of size smaller than $1/m$. It can be reliably calculated in the dilute gas approximation.

ACKNOWLEDGMENTS

I am grateful to Steven Weinberg for many valuable suggestions and to Sam Osofsky for discussions of the material in Section 4.5.

Literature Cited

1. Georgi H. *Weak Interactions and Modern Particle Theory*. Menlo Park, Calif: Benjamin/Cummings (1984)
2. Isgur N, Wise MB. *Phys. Lett.* 232B: 113 (1989); 208B:504 (1988); Eichten E, Hill B. *Phys. Lett.* 234B:511 (1990); Grinstein B. *Nucl. Phys.* B339:253 (1990); Georgi H. *Phys. Lett.* B240:447 (1990)

2a. Grinstein B. *Annu. Rev. Nucl. Part. Sci.* 42:101–45 (1992)

3. Weinberg S. *Phys. Rev. Lett.* 18:188 (1967); Dashen R. *Phys. Rev.* 183:1245 (1969); Dashen R, Weinstein M. *Phys. Rev.* 183:1291 (1969)
4. Weinberg S. *Physica* A96:327 (1979)
5. Li LF, Pagels H. *Phys. Rev. Lett.* 26: 1204 (1971); see also Pagels H. *Phys. Rep.* 16C:531 (1976) and references therein
6. Gasser J, Leutwyler H. *Ann. Phys. NY* 158:142 (1984); *Nucl. Phys.* B250:465, 517, 539 (1985)
7. Coleman S, Wess J, Zumino B. *Phys. Rev.* 177:2239 (1969); Callan CG, et al. *Phys. Rev.* 177:2247

7a. Weinberg S. *Phys. Rev.* 166:1568 (1968)

8. t Hooft G. *Nucl. Phys.* B33:173 (1971)
9. Glashow SL. *Nucl. Phys.* B22:579 (1961)
10. Weinberg S. *Phys. Rev. Lett.* 19:1264 (1967)
11. Salam A. In *Elementary Particle Theory: Relativistic Groups and Analyticity* (Nobel Symp. No. 8), ed. N Svartholm. New York: Wiley (1969)
12. Georgi H, Glashow SL. *Phys. Rev. Lett.* 32:438 (1974); Pati J, Salam A. *Phys. Rev.* D10:275 (1974)
13. Gildener E, Weinberg S. *Phys. Rev.* D13:3333 (1976); see also Ovrut B, Schnitzer HJ. *Phys. Rev.* D21:3369 (1980)
14. Glashow SL, Weinberg S. *Phys. Rev.* D15:1958 (1977)
15. Georgi H, Nanopoulos DV. *Phys. Lett.* 82B:95 (1979)
16. Glashow SL, Iliopoulos J, Maiani L. *Phys. Rev.* D2:1285 (1970)
17. Appelquist T, Carazone J. *Phys. Rev.* D11:2856 (1975)
18. Georgi H, Quinn H, Weinberg S. *Phys. Rev. Lett.* 33:451 (1974)
19. Georgi H, Politzer HD. *Phys. Rev.* D14:1829 (1976)
20. Ross DA. *Nucl. Phys.* B140:1 (1978); Goldman TJ, Ross DA. *Phys. Lett.* B84:208 (1979)
21. Georgi H, Politzer HD. *Nucl. Phys.* B117:1829 (1976)
22. Witten E. *Nucl. Phys.* B122:109 (1977)
23. Weinberg S. *Phys. Lett.* B91:51 (1980); see also Ovrut B, Schnitzer HJ. *Phys. Lett.* B100:2518 (1981)
24. Hall L. *Nucl. Phys.* B178:75 (1981)
25. Witten E. *Nucl. Phys.* B104:445 (1976)
26. Gilman FJ, Wise MB. *Phys. Rev.* D20: 2392 (1979)
27. Lee B, Gaillard MK. *Phys. Rev. Lett.* 13:108 (1974)
28. Altarelli G, Maiani L. *Phys. Lett.* B52: 351 (1974)
29. Ovrut B, Schnitzer HJ. *Phys. Rev.* D24:1695 (1982)
30. Arst C. Reduced effective lagrangians, *Michigan Preprint M-TH-92-28* (1992)
31. Fritzsch H, Gell-Mann M. In *Proc. 16th Int. Conf. on High Energy Physics,* Chicago (1972), Vol. 2; Weinberg S. *Phys. Rev. Lett.* 31:494 (1973)
32. Wilczek F. *Phys. Rev.* D29:338 (1984); see also Rajagopal K, Wilczek F. *Princeton Univ. preprint PUPT-1347* (1993), to be published
33. Georgi H, Kaplan D. *Phys. Lett.* B136: 183 (1984); see also Dugan M, Georgi

H, Kaplan D. *Nucl. Phys.* B254:299 (1985) and references therein
34. Sikivie P, et al. *Nucl. Phys.* B173:189 (1980)
35. Miransky VA, Tanabashi M, Yamawaki K. *Phys. Lett.* B221:177 (1989); *Mod. Phys. Lett.* A4:1043 (1989); Nambu Y. *Univ. Chicago preprint EFI 89-08* (Feb. 1989); Marciano WJ. *Phys. Rev. Lett.* 62:2793 (1989)
36. Hill C. *Phys. Lett.* B226:419 (1991)
37. Bardeen WA, Hill CT, Lindner M. *Phys. Rev.* D41:1647 (1990)
38. Manohar A, Georgi H. *Nucl. Phys.* B234:189 (1984); Georgi H, Randall L. *Nucl. Phys.* B276:241 (1986); see also Espriu D, De Rafael E, Taron J. *Nucl. Phys.* B345:22 (1990)
39. Lee BW, Quigg C, Thacker HB. *Phys. Rev. Lett.* 38:883 (1977)
40. Farhi E, Susskind L. *Phys. Rep.* 74:277 (1981) and references therein
41. Deleted in proof
42. Chanowitz M, Golden M, Georgi H. *Phys. Rev.* D36:1490 (1987)
43. Georgi H. *Nucl. Phys.* B331:311 (1990)
44. Feruglio F. *Lectures Given at the 2nd National Seminar on Theoretical Physics,* Parma, 1–12 Sept. (1992) and references therein
45. Peskin ME, Takeuchi T. *Phys. Rev. Lett.* 65:964 (1990)
46. Golden M, Randall L. *Nucl. Phys.* B361:3 (1991)
47. Holdom B, Terning J. *Phys. Lett.* B247:88 (1990)
48. Grinstein B, Wise MB. *Phys. Lett.* B265:326 (1991)
49. Jenkins E, Manohar A. *Phys. Lett.* B237:259 (1989)
50. Cohen A, Georgi H, Grinstein B. *Nucl. Phys.* B232:61 (1984)
51. Lynn BW, Peskin ME, Stuart RG, *SLAC-PUB-3725* (1985); Kennedy DC, Lynn BW. *Nucl. Phys.* B322:1 (1989)
52. Kuroda M, Moultaka G, Schildknecht D. *Nucl. Phys.* B350:25 (1991)
53. Consoli M, et al. *Phys. Lett.* 227B:167 (1989)
54. Gross D, Andrei N. *Phys. Rev.* D18:468 (1978) and references therein