EFFECTIVE GAUGE THEORIES ☆

Steven WEINBERG
Lyman Laboratory of Physics, Harvard University, Cambridge, MA 02138, USA
and Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

Received 7 January 1980

A method is described for dealing with gauge field theories that contain very heavy particles, by constructing a gauge-invariant effective field theory, in which the heavy particles do not appear. For theories in which a simple group is spontaneously broken to the strong and electroweak gauge groups, the running strong and electroweak effective couplings become equal to the original gauge coupling at different renormalization scales, given as weighted geometric averages of the superheavy particle masses. The method is also applied to quantum chromodynamics with heavy quarks.

If weak, electromagnetic, and strong couplings become comparable at any energy, it is likely to be at an energy enormously greater than those with which we are generally familiar [1]. Specifically, if a large simple gauge group G ‡1 is spontaneously broken at a scale M to the SU(3) and SU(2) × U(1) gauge groups of the strong and electroweak interactions (plus possible other subgroups of G which commute with SU(3) × SU(2) × U(1)), and if all fermions are essentially similar to observed quarks and leptons (plus possible SU(3) × SU(2) × U(1) neutrals), then the mass scale M of the superheavy gauge bosons is given by ‡2

$$\ln \frac{M}{m} = \tfrac{4}{11}\pi^2 \left[\frac{1}{e^2(m)} - \frac{8}{3g_S^2(m)} \right] + \mathrm{O}(1) \simeq 35 \ , \qquad (1)$$

and the Z^0–γ mixing parameter is ‡2

$$\sin^2\theta = \tfrac{1}{6} + \tfrac{5}{9} e^2(m)/g_S^2(m) + \mathrm{O}(\alpha) \simeq 0.2 \ , \qquad (2)$$

where $g_S(m)$ and $e(m)$ are the strong and electromagnetic couplings measured at some "ordinary" mass scale m, say $m \approx 100$ GeV.

☆ Research is supported in part by the National Science Foundation under Grant No. PHY77-22864.

‡1 Specific examples of such simple groups are given in the models of ref. [2].

‡2 See ref. [1]. Small effects of scalar particles are neglected in eqs. (1) and (2).

Clearly, in order to calculate the superheavy gauge boson masses with any precision it is necessary to calculate the O(1) term in eq. (1). Also, experimental determinations of $\sin^2\theta$ are approaching the point where we would like to know the $\mathrm{O}(\alpha)$ term in eq. (2). But to go beyond the lowest order results (1), (2), care is needed in the definition of the running couplings $g_i(\mu)$, and in their determination from experimental data at $\mu \approx m$.

Such calculations have recently been presented by Goldman and Ross and by Marciano [3]. In their method, the running SU(3) × SU(2) × U(1) couplings $g_i(\mu)$ are defined as the values of various off-shell Green's functions at renormalization points with momentum scale μ, and the Appelquist–Carrazone theorem [4] is invoked to justify the neglect of superheavy particles at ordinary energies. There is nothing in principle that is wrong with this approach, but it has some awkward features. In particular, in the method of ref. [3] it is not possible to take advantage of the calculational simplicity of the "minimum subtraction" definition [5] of renormalized coupling constants (or other mass-independent renormalization schemes [6]), because with renormalizations carried out by minimum subtraction, the Appelquist–Carrazone theorem would not hold [7]. It is also complicated to include the superheavy gauge bosons in the renormalization group equations for $g_i(\mu)$, which have to be evaluated to two-loop order in order to determine the O(1) terms in eq. (1) and the $O(\alpha)$ terms in eq. (2).

In this note, I would like to lay the groundwork for a different approach to this sort of calculation. The method is based on the use of an SU(3) × SU(2) × U(1)-invariant "effective field theory", in which all superheavy fields have been integrated out, and the only fields that appear explicitly are those with ordinary masses, much less than M.

In general, if a field theory contains a set of fields ϕ of small or zero mass, and another set Φ of much larger mass, then the action $\tilde{I}[\phi]$ of the effective field theory may be obtained from the action $I[\phi, \Phi]$ of the full field theory by a functional integral over Φ,

$$\exp(i\tilde{I}[\phi]) = \int [d\Phi] \exp(iI[\phi, \Phi]) . \tag{3}$$

The expectation value of any functional of the ϕ fields can be calculated as an integral over ϕ with weight $\exp(i\tilde{I})$, or as an integral over ϕ and Φ with weight $\exp(iI)$. Of course, $\tilde{I}$ will contain an infinite number of nonrenormalizable interactions, and the effective field theory cannot conveniently be used in calculations of physical processes at energies comparable with the masses of the Φ fields. However, any nonrenormalizable coupling constant in $\tilde{I}$ that has dimensionality $[\text{mass}]^d$ with $d < 0$ may be expected to be roughly of order M^d, where M is a typical mass of the Φ fields. Hence the effects of the nonrenormalizable couplings are suppressed [‡3] at energies $E \ll M$ by powers of E/M. Since there are no superheavy particles in the effective field theory, the Appelquist–Carrazone theorem is not needed, and there is nothing to prevent the use of the minimum subtraction definition of renormalized coupling constants. Indeed, this definition has a feature that adds greatly to the simplicity of our approach: the renormalization-group equations for the renormalizable couplings in $\tilde{I}$ decouple from the others [‡4], so that we can ignore all the nonrenormalizable couplings in working out the variation of the running renormalizable couplings with energy scale μ, even for μ comparable with M!

These remarks lead us to a program for calculating the O(1) terms in $\ln(M/m)$ and the $O(\alpha)$ terms in $\sin^2\theta$, as follows:

(a) Determine the renormalized SU(3) × SU(2) × U(1) couplings $g_i(\mu)$ in $\tilde{I}$ at renormalization scales μ of order M, by a *one-loop* functional integral over superheavy fields, as in (3). (Here and below, all renormalized couplings are to be defined by minimum subtraction [5], or by one of its simple modifications [10].)

(b) Calculate the renormalized SU(3) × SU(2) × U(1) couplings $g_i(\mu)$ at ordinary mass scales $\mu \approx m$, by integrating the *two-loop* renormalization group equations of the SU(3) × SU(2) × U(1) gauge theory (in which no superheavy fields appear) from $\mu \approx M$ to $\mu \approx m$, using the results of (a) as an initial condition at $\mu \approx M$.

(c) Compare the results of (b) with the SU(3) × SU(2) × U(1) couplings determined from experiment at ordinary energy, including the effects of radiative corrections to *one-loop* order.

There is a difficulty in step (a), having to do with gauge invariance. In the spontaneous breakdown of the simple group G, a gauge subgroup $\tilde{G}$, is left unbroken (presumably, $\tilde{G}$ is SU(3) × SU(2) × U(1), perhaps with extra factors), and we would like the effective action $\tilde{I}$ to be gauge invariant under $\tilde{G}$. However, in order to calculate the functional integral over the superheavy gauge fields in (3), it is necessary to add a gauge-fixing term to $I[\phi, \Phi]$, and such terms usually spoil gauge invariance under the unbroken subgroup $\tilde{G}$ as well as under the rest of G. For instance, one convenient gauge [11] is specified by adding to I a term:

$$I_{\text{gauge fix}} = -\frac{1}{2} \sum_\alpha \int d^4x\, f_\alpha^2(x; \phi, \Phi) , \tag{4}$$

where

$$f_\alpha = \xi^{-1/2} [\partial_\mu V_\alpha^\mu + ig\xi(\lambda, t_{\alpha S} S)] . \tag{5}$$

‡3 Some of these suppressed interactions may become observable if they violate otherwise exact conservation laws. For a discussion of baryon and lepton nonconservation from this point of view, see ref. [8].

‡4 With renormalized couplings defined by minimum subtraction, the logarithmic derivative of a running coupling $g_n(\mu)$ can contain a term of order $g_m(\mu)g_l(\mu)$ ... only if the dimensionality of g_n (in powers of mass) equals the sum of the dimensionalities of g_m, g_l, etc. If there are no masses or superrenormalizable couplings, then all couplings have negative or zero dimensionality, so the logarithmic derivative of a renormalizable coupling of zero dimensionality can only involve other couplings of zero dimensionality. (See e.g. ref. [9]). Effects of small masses are suppressed by powers of the ratio of these masses to M.

In the notation to be used here, α runs over all generators of the full gauge group G, V_α^μ are the corresponding gauge fields; S is a column of hermitian scalar fields; $t_{\alpha S}$ is the matrix representing the αth generator of G on these scalars; λ is the column of zeroth-order vacuum expectation values of S; g is the unrenormalized gauge coupling constant of G; and ξ is a free parameter characterizing the choice of gauge. It will also be convenient to adopt a notation in which lower case Latin letters $a, b, c, \ldots$ represent values of α corresponding to the generators of the unbroken subgroup $\widetilde{G}$ of G, while upper case Latin letters $A, B, C, \ldots$ are values of α corresponding to the other generators of G. (The gauge fields V_a^μ of $\widetilde{G}$ are thus included among the light fields ϕ, while the other gauge fields V_A^μ are included among the superheavy fields Φ.) The sum in eq. (4) then splits into a sum over "unbroken" indices a, plus another sum over "broken" indices A. Now, there is no difficulty with the first sum: the unbroken generators satisfy $\theta_a\lambda = 0$, so $f_a = \partial_\mu V_a^\mu/\sqrt{\xi}$, and since this is independent of all superheavy fields, the term $-\frac{1}{2}\int\Sigma_\alpha f_a^2\, d^4x$ in (4) can be ignored for the present, and brought back later as a $\widetilde{G}$-gauge fixing term when we do functional integrals over the ϕ fields. That is, in place of (4), we may fix the gauge of the superheavy V_A^μ fields by adding to I only a term

$$\Delta I = -\frac{1}{2}\int d^4x \sum_A f_A^2(X;\phi,\Phi)\,. \tag{6}$$

The difficulty is that although ΔI is globally invariant under the unbroken subgroup $\widetilde{G}$, it is not locally invariant, because the superheavy gauge fields V_A^μ transform nontrivially under $\widetilde{G}$, and an ordinary derivative acts on these fields in ΔI.

Once the problem is stated in this way, the solution is obvious: replace the derivative in (5) with a $\widetilde{G}$ covariant derivative. That is, in place of (5), take the gauge-fixing functions as [‡5]

$$f_A = \xi^{-1/2}[\partial_\mu V_A^\mu + gC_{ABa}V_B^\mu V_{a\mu} + ig\xi(\lambda, t_{AS}S)]\,, \tag{7}$$

with $C_{\alpha\beta\gamma}$ the structure constant of G [‡6]. This is $\widetilde{G}$-gauge covariant, in the sense that under $\widetilde{G}$-gauge transformation $\phi\to\phi_\Omega$, $\Phi\to\Phi_\Omega$ with infinitesimal parameters $\Omega_a(X)$, the functions f_A undergo a linear transformation

$$\left[\frac{\delta f_A(X;\phi_\Omega,\Phi_\Omega)}{\delta\Omega_a(y)}\right]_{\Omega=0} = -i\delta^4(x-y)C_{ABa}f_B(x;\phi,\Phi)\,. \tag{8}$$

Hence (6) is now $\widetilde{G}$-gauge-invariant. Our prescription for the effective action $\widetilde{I}$ is then

$$\exp(i\widetilde{I}[\phi]) = \int d\Phi\,\exp(iI[\phi,\Phi] + i\Delta I[\phi,\Phi]) \times \mathrm{Det}\,X[\phi,\Phi]\,, \tag{9}$$

where ΔI is given by eqs. (6) and (7), and X is the "matrix"

$$X_{Ax,By} = [\delta f_A(x;\phi_\Omega,\Phi_\Omega)/\delta\Omega_B(y)]_{\Omega=0}\,. \tag{10}$$

This prescription manifestly gives an effective action $\widetilde{I}$ which is gauge invariant under $\widetilde{G}$. But it is not immediately obvious that the prescription is correct. To see the problem, suppose we use the effective field theory to calculate S-matrix elements or expectation values of gauge-invariant functionals of ϕ. We would insert a $\widetilde{G}$-gauge-fixing functional

$$\exp\left[-\frac{1}{2}\int\sum_a f_a^2(x;\phi)\,d^4x\right],$$

and a corresponding determinantal factor Det $Y[\phi]$, with $Y_{ax,by}$ given by a formula like (10), but with unbroken indices a, b in place of the broken indices A, B. If we use (9) to express $\widetilde{I}$ in terms of the original action I, we find just the usual Faddeev–Popov–de Witt formula [13] except that Det X Det Y appears in place of the determinant of the full matrix $Z_{\alpha x,\beta y}$, which is defined as in eq. (10) but with general indices α, β in place of the broken indices A, B. Fortunately this difference does not matter. It is well known that the gaussian functional (4) can be replaced with a product of delta functions of $f_\alpha(x)$ in the usual FPdW formula [13], and it can easily be seen that the same is true with Det X Det Y in place of Det Z. Thus we may start with the usual FPdW formula, replace the gaussian (4) with a product of delta functions, use eq. (8) and the fact that now $f_\alpha = 0$ to show that the off-diagonal block $Z_{Ax,ay}$ vanishes, so that Det Z = Det X Det Y, and then change the product of delta functions back to a gaussian. But even though the final results are the same, it is very convenient to have Det X Det Y in place

‡5 This is formally very similar to the introduction of a background field gauge in ref. [12].

‡6 The generators of G are normalized so that the structure constants $C_{\alpha\beta\gamma}$ are totally antisymmetric.

of Det Z, because only in this way can we construct a $\widetilde{G}$ gauge invariant effective action $\widetilde{\Gamma}$.

This prescription has been used to calculate the effective gauge coupling constants in a very general context [‡7]. We consider an arbitrary renormalizable gauge theory, with a simple gauge group G, and suppose that G is spontaneously broken to a direct product $\widetilde{G}$ of simple (or U(1)) subgroups $\widetilde{G}_i$, giving superlarge masses to various gauge bosons, scalars, and perhaps spin-$\frac{1}{2}$ fermions. The unrenormalized couplings g_i that are associated in $\widetilde{\Gamma}$ with these subgroups can be determined from the part of $\widetilde{\Gamma}$ that is quadratic in the Yang–Mills curls $F_{a\mu\nu}$, which (since $\widetilde{\Gamma}$ is gauge invariant) is given by the part of $\widetilde{\Gamma}$ quadratic in the ordinary curls $\partial_\mu V_{a\nu} - \partial_\nu V_{a\mu}$. This can be evaluated by calculating the graphs with two external $V_{a\mu}$ lines and only superheavy internal lines. In one-loop order, we find [‡7]

$$g_i = g + \frac{g^3\Gamma(2-D/2)}{12(4\pi)^{D/2}}\,[-\mathrm{Tr}(t_{iS}^2 M_S^{D-4}\Lambda) - 2^{1+D/2}\mathrm{Tr}(t_{iF}^2 M_F^{D-4}) + (25-D)\mathrm{Tr}(t_{iV}^2 M_V^{D-4})]\,. \tag{11}$$

Here g is the unrenormalized gauge coupling of the original theory and t_{iS}, t_{iF}, and t_{iV} are the matrices which represent any one of the generators of $\widetilde{G}_i$ on the superheavy scalars, spin-$\frac{1}{2}$ fermions, or vectors, respectively [‡6]. The traces are to be taken over all representations of $\widetilde{G}$ containing superheavy particles of spin 0, $\frac{1}{2}$, 1; Λ is a projection operator which excludes the Goldstone bosons in these representations; and M_S, M_F, and M_V are the mass matrices of the superheavy particles of spin 0, $\frac{1}{2}$, and 1. Dimensional regularization [13] has been used, with D the dimensionality of space–time.

The "modified minimum subtraction" renormalized couplings [10] $g_i(\mu)$ and $g(\mu)$ are defined as the remainder terms in $g_i\mu^{D/2-2}$ and $g\mu^{D/2-2}$, after separating out the pole terms proportional to $(D-4)^{-1} + \gamma/2 - \ln\sqrt{4\pi}$, where $\gamma = 0.577\ \ldots$. Comparing these remainder terms in (11), we find

$$g_i(\mu) = g(\mu) + \frac{g(\mu)^3}{96\pi^2}\{\mathrm{Tr}[t_{iS}^2\Lambda\ln(M_S/\mu)] + 8\,\mathrm{Tr}[t_{iF}^2\ln(\sqrt{2}M_F/\mu)] + \mathrm{Tr}\,t_{iV}^2 - 21\,\mathrm{Tr}[t_{iV}^2\ln(M_V/\mu)]\}\,. \tag{12}$$

‡7 Details of the calculation will be given in a paper now in preparation.

The only approximation which has been made in deriving eq. (12) is to neglect terms of higher order in $g(\mu)$. This approximation is presumably valid as long as the logarithms are not too large, i.e., for μ roughly comparable with M_S, M_F, and M_V. For such μ's, it can be checked that the μ-dependence of $g_i(\mu)$ given by (12) is the same as would be calculated directly in the effective $\widetilde{G}$-gauge theory. Of course, for $\mu \ll M$, we must use the renormalization group equations of the effective field theory rather than (12) to calculate the $g_i(\mu)$, with (12) serving as the initial condition at $\mu \approx M$.

Eq. (12) shows that the individual running couplings $g_i(\mu)$ become equal to the coupling $g(\mu)$ of the original theory at different points μ_i, given as suitably weighted geometric averages of superheavy masses, with the weighting determined by the $\widetilde{G}_i$ quantum numbers of these superheavy particles. The factor 21 in the last term of (12) is sufficiently large so that it should be the superheavy vector boson masses that chiefly determine the μ_i, especially if there are no superheavy fermions.

For an illustration, let us suppose that the spectrum of superheavy gauge bosons is the same as in the O(10) model [‡8]: labelling them by SU(3) multiplicity, SU(2) multiplicity, and weak hypercharge, the five superheavy gauge multiplets are A(3,2,5/6), B($\overline{3}$,2,1/6), C($\overline{3}$,1,2/3), D(1,1,1), and E(1,1,0). Assuming that the last term in (12) is dominant, the mass scales where the SU(3), SU(2), and U(1) couplings become equal to $g(\mu)$ are given by

$$\mu_3 = m_A^{2/5} m_B^{2/5} m_C^{1/5}\,,\quad \mu_2 = m_A^{1/2} m_B^{1/2}\,,$$
$$\mu_1 = m_A^{5/8} m_B^{1/40} m_C^{1/5} m_D^{3/20}\,.$$

The methods of this paper can also be useful in dealing with heavy quarks in quantum chromodynamics (QCD). Consider QCD with $N-1$ light quarks and one heavy quark of mass M, and construct an effective field theory by integrating out the heavy quark. Eq. (12) shows that the gauge coupling of the effective theory is equal to the gauge coupling of the original theory at $\mu = \sqrt{2}\,M$ [‡9]. By equating couplings at $\sqrt{2}\,M$, one can calculate the two-loop QCD scale factor Λ_N of the

‡8 See the articles by Georgi, Fritzsch, Minkowski, and Nanopoulos in ref. [2].

‡9 This result has also been obtained in unpublished calculations by L. Hall.

N-quark theory in terms of the corresponding quantity Λ_{N-1} in the $N-1$-quark effective theory.

For instance, for $N = 5$ and $M = m_b = 5$ GeV, if $\Lambda_4 = 400$ MeV, then $\Lambda_5 = 270$ MeV.

Added note. The construction of effective field theories of various sorts has been considered recently by B. Ovrut and H.J. Schnitzer, Brandeis preprint; Y. Kazama and Y.P. Yao, Michigan preprint UM HE 79-40; T. Hagiwara and N. Nakazawa, paper in preparation.

I am grateful for valuable conversations with H. Georgi.

References

[1] H. Georgi, H.R. Quinn and S. Weinberg, Phys. Rev. Lett. 33 (1974) 451;
also see: A. Buras, J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Nucl. Phys. B135 (1978) 66.

[2] J.C. Pati and A. Salam, Phys. Rev. D10 (1974) 1240;
H. Georgi and S.L. Glashow, Phys. Rev. Lett. 32 (1974) 438;
H. Georgi, in: Particles and fields (American Institute of Physics, New York, 1975);
H. Fritzsch and P. Minkowski, Ann. Phys. 93 (1975) 183;
H. Georgi and D.V. Nanopoulos, Phys. Lett. 82B (1979) 392;
F. Gürsey, P. Ramond and P. Sikivie, Phys. Lett. 60B (1975) 177;
F. Gürsey and P. Sikivie, Phys. Rev. Lett. 36 (1976) 775;
P. Ramond, Nucl. Phys. B110 (1976) 214; etc.

[3] D. Ross, Nucl. Phys. B140 (1978) 1;
W.J. Marciano, Phys. Rev. D20 (1979) 274;
T. Goldman and D. Ross, CALT 68-704, to be published.

[4] T. Appelquist and J. Carrazzone, Phys. Rev. 11 (1975) 2856.

[5] G. 't Hooft, Nucl. Phys. B61 (1973) 455; B62 (1973) 444.

[6] S. Weinberg, Phys. Rev. D8 (1973) 3497.

[7] See e.g.: P. Binétruy and T. Schücker, paper in preparation.

[8] S. Weinberg, Phys. Rev. Lett. 43 (1979) 1566;
F. Wilczek and A. Zee, Phys. Rev. Lett. 43 (1979) 1571.

[9] S. Weinberg, in: General relativity; an Einstein centenary, eds. S.W. Hawking and W. Israel (Cambridge. U.P. 1979) p. 817.

[10] W.A. Bardeen, A.J. Buras, D.W. Duke and T. Muta, Phys. Rev. D18 (1978) 3998.

[11] K. Fujikawa, B.W. Lee and A.I. Sanda, Phys. Rev. D6 (1972) 2923.

[12] B.S. de Witt, Phys. Rev. 162 (1967) 1195, 1239; Phys. Rep. 19 (1975) 295;
G. 't Hooft and M. Veltman, Ann. Inst. H. Poincaré 20 (1974) 69;
S.J. Honerkamp, Nucl. Phys. B48 (1972) 269;
R. Kallosh, Nucl. Phys. B78 (1974) 293;
M.T. Grisaru, P. van Nieuwenhuizen and C.C. Wu, Phys. Rev. D12 (1975) 3203.

[13] L.D. Fadeev and V.N. Popov, Phys. Lett. 25B (1967) 29;
B. deWitt, Phys. Rev. 162 (1967) 1195, 1239.