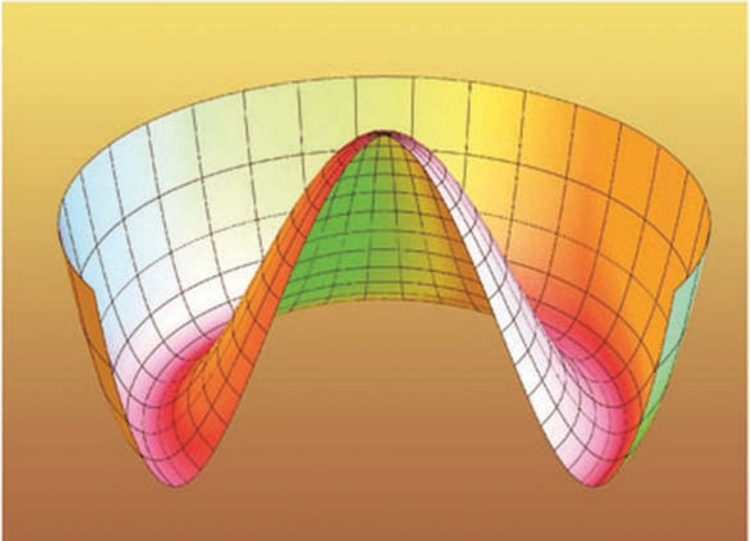
Paul H. Frampton

WILEY-VCH

# Gauge Field Theories

Third, Enlarged and Improved Edition



*Paul H. Frampton*

**Gauge Field Theories**

## *Related Titles*

Reiser, M.

**Theory and Design of Charged Particle Beams**

**634 pages with 107 figures**

1994. Hardcover

ISBN: 978-0-471-30616-0

Edwards, D. A., Syphers, M. J.

**An Introduction to the Physics of High Energy Accelerators**

**304 pages with 103 figures**

1993. Hardcover

ISBN: 978-0-471-55163-8

Griffiths, D.

**Introduction to Elementary Particles**

**450 pages with 150 figures**

2008. Softcover

ISBN: 978-3-527-40601-2

*Paul H. Frampton*

# Gauge Field Theories

3rd, Enlarged and Improved Edition

WILEY-VCH

WILEY-VCH Verlag GmbH & Co. KGaA

**The Author**

***Prof. Paul H. Frampton***
University of North Carolina
Dept. of Physics and Astronomy
Chapel Hill, USA
frampton@physics.unc.edu

**Cover Picture**
Plot of the function $V(x, y) = -a(x^2 + y^2) + b(x^2 + y^2)^2$ (created with MATHEMATICA 6 by R. Tiebel)



**Library of Congress Card No.:**
applied for

**British Library Cataloguing-in-Publication Data**
A catalogue record for this book is available from the British Library.

**Bibliographic information published by the Deutsche Nationalbibliothek**
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.



**Typesetting** VTEX, Litauen

**Printing** Strauss GmbH, Mörlenbach

**Binding** Litges & Dopf buchbinderei GmbH, Heppenheim

Printed in the Federal Republic of Germany
Printed on acid-free paper

**ISBN** 978-3-527-40835-1

# Contents

## Preface to the Third Edition

The first edition of *Gauge Field Theories* was published in 1986, the second in 2000. In summer and fall of 2007, Christoph von Friedeburg from Wiley-VCH decided to persuade me to prepare this third edition anticipating the commissioning of the Large Hadron Collider (LHC) at CERN, near Geneva in Switzerland.

Whatever the LHC finds, essentially all the material on the standard model in the first six chapters of *Gauge Field Theories* will remain valid. The only exception is that, if the Higgs boson is found, the newly modified Section 4.7 in Chapter 4 could then be turned into an entire book.

Unlike in 1986, there are now many other excellent texts on the same subject. Therefore it seemed allowed to make the completely revised 25% at the end cover different ground. The last Chapters 7 and 8 of *Gauge Field Theories* are now concerned with the topic of model building. At most one, and more likely none, of the specific models discussed there will be of permanent value. In the final chapter of GFT I discuss the relevance to model building of four dimensional conformal invariance which other books sometimes omit.

It is pleasing that *Gauge Field Theories* continues to sell although, as mentioned in the preface to the second edition, the biggest reward for the effort involved in writing GFT is when someone tells me he or she received inspiration from reading it.

Finally, I wish to thank Christoph von Friedeburg and Ulrike Werner at Wiley for their encouragement to prepare a third edition.

Chapel Hill, February 2008 *Paul H. Frampton*

## Preface to the Second Edition

The writing of the first edition of *Gauge Field Theories* was spread over ten years. The first chapter was written in 1976 in connection with some lectures at UCLA, but the book was completed only a decade later in Chapel Hill.

It has been gratifying to learn that the book has been used as a basis for graduate courses in universities and other institutions of higher learning in several different countries. But it was even more rewarding when just one physicist would tell me that he or she had learned gauge theory from it. Luckily, there has been more than one such experience since the first edition appeared.

In late 1998 an editor from Wiley convinced me to undertake this second edition. In preparing it, I have updated every chapter. The material on gauge invariance, quantization, and renormalization in Chapters 1 to 3 has been rearranged in a more logical order. Most of Chapter 4 on electroweak interactions has been rewritten entirely to accommodate the latest precision data and the discovery of the top quark. Chapters 5 and 6 cover the renormalization group and quantum chromodynamics. In Chapter 7, which is entirely new, model building is discussed.

My hope is that at least one student aspiring to do research in theoretical physics will tell me that he or she learned field theory from the second edition of the book.

Chapel Hill, November 1999 *P.H. Frampton*

# 1
# Gauge Invariance

## 1.1
## Introduction

Gauge field theories have revolutionized our understanding of elementary particle interactions during the second half of the twentieth century. There is now in place a satisfactory theory of strong and electroweak interactions of quarks and leptons at energies accessible to particle accelerators at least prior to LHC.

All research in particle phenomenology must build on this framework. The purpose of this book is to help any aspiring physicist acquire the knowledge necessary to explore extensions of the standard model and make predictions motivated by shortcomings of the theory, such as the large number of arbitrary parameters, and testable by future experiments.

Here we introduce some of the basic ideas of gauge field theories, as a starting point for later discussions. After outlining the relationship between symmetries of the Lagrangian and conservation laws, we first introduce global gauge symmetries and then local gauge symmetries. In particular, the general method of extending global to local gauge invariance is explained.

For global gauge invariance, spontaneous symmetry breaking gives rise to massless scalar Nambu–Goldstone bosons. With local gauge invariance, these unwanted particles are avoided, and some or all of the gauge particles acquire mass. The simplest way of inducing spontaneous breakdown is to introduce scalar Higgs fields by hand into the Lagrangian.

## 1.2
## Symmetries and Conservation Laws

A quantum field theory is conveniently expressed in a Lagrangian formulation. The Lagrangian, unlike the Hamiltonian, is a Lorentz scalar. Further, important conservation laws follow easily from the symmetries of the Lagrangian density, through the use of Noether's theorem, which is our first topic. (An account of Noether's theorem can be found in textbooks on quantum field theory, e.g., Refs. [1] and [2].)

Later we shall become aware of certain subtleties concerning the straightforward treatment given here. We begin with a Lagrangian density

$$\mathscr{L}\big(\phi_k(x), \partial_\mu \phi_k(x)\big) \tag{1.1}$$

where $\phi_k(x)$ represents genetically all the local fields in the theory that may be of arbitrary spin. The Lagrangian $L(t)$ and the action $S$ are given, respectively, by

$$L(t) = \int d^3x \mathscr{L}\big(\phi_k(x), \partial_k \phi_k(x)\big) \tag{1.2}$$

and

$$s \int_{t_1}^{t_2} dt L(t) \tag{1.3}$$

The equations of motion follow from the Hamiltonian principle of stationary action,

$$\delta S = \delta \int_{t_1}^{t_2} dt\, d^3x \mathscr{L}\big(\phi_k(x), \partial_\mu \phi_k(x)\big) \tag{1.4}$$

$$= 0 \tag{1.5}$$

where the field variations vanish at times $t_1$ and $t_2$ which may themselves be chosen arbitrarily.

It follows that (with repeated indices summed)

$$0 = \int_{t_1}^{t_2} dt\, d^3x \left[ \frac{\partial \mathscr{L}}{\partial \phi_k} \delta\phi_k + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)} \delta(\partial_\mu \phi_k) \right] \tag{1.6}$$

$$= \int_{t_1}^{t_2} dt\, d^3x \left[ \frac{\partial \mathscr{L}}{\partial \phi_k} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)} \right] \delta\phi_k + \left[ \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)} \delta\phi_k \right]_{t=t_1}^{t=t_2} \tag{1.7}$$

and hence

$$\frac{\partial \mathscr{L}}{\partial \phi_k} = \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)} \tag{1.8}$$

which are the Euler–Lagrange equations of motion. These equations are Lorentz invariant if and only if the Lagrangian density $\mathscr{L}$ is a Lorentz scalar.

The statement of Noether's theorem is that to every continuous symmetry of the Lagrangian there corresponds a conservation law. Before discussing internal symmetries we recall the treatment of symmetry under translations and rotations.

Since $\mathscr{L}$ has no explicit dependence on the space–time coordinate [only an implicit dependence through $\phi_k(x)$], it follows that there is invariance under the translation

$$x_\mu \to x'_\mu = x_\mu + a_\mu \tag{1.9}$$

where $a_\mu$ is a four-vector. The corresponding variations in $\mathscr{L}$ and $\phi_k(x)$ are

$$\delta\mathscr{L} = a_\mu \partial_\mu \mathscr{L} \tag{1.10}$$

$$\delta\phi_k(x) = a_\mu \partial_\mu \phi_k(x) \tag{1.11}$$

Using the equations of motion, one finds that

$$a_\mu \partial_\mu \mathscr{L} = \frac{\partial \mathscr{L}}{\partial \phi_k}\delta\phi_k + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)}\delta(\partial_\mu \phi_k) \tag{1.12}$$

$$= \partial_\mu \left[ \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)}\delta\phi_k \right] \tag{1.13}$$

$$= a_\nu \partial_\mu \left[ \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)}\partial_\nu\phi_k \right] \tag{1.14}$$

If we define the tensor

$$T_{\mu\nu} = -g_{\mu\nu}\mathscr{L} + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_k)}\partial_\nu \phi_k \tag{1.15}$$

it follows that

$$\partial_\mu T_{\mu\nu} = 0 \tag{1.16}$$

This enables us to identify the four-momentum density as

$$\mathscr{P}_\mu = T_{0\mu} \tag{1.17}$$

The integrated quantity is given by

$$P_\mu = \int d^3x\, \mathscr{P}_\mu \tag{1.18}$$

$$= \int d^3x(-g_{0\mu}\mathscr{L} + \pi_k \partial_\mu \phi_k) \tag{1.19}$$

where $\pi_k = \partial\mathscr{L}/\partial\dot{\phi}_k$ is the momentum conjugate to $\phi_k$. Notice that the time component is

$$\mathscr{P}_0 = \pi_k \partial_0 \phi_k - \mathscr{L} \tag{1.20}$$

$$= \mathscr{H} \tag{1.21}$$

where $\mathscr{H}$ is the Hamiltonian density. Conservation of linear momentum follows since

$$\frac{\partial}{\partial t} P_\mu = 0 \tag{1.22}$$

This follows from $P_i = J_{0i}$ and $\frac{\partial}{\partial t} J_{0i}$ becomes a divergence that vanishes after integration $\int d^3x$.

Next we consider an infinitesimal Lorentz transformation

$$x_\mu \to x'_\mu = x_\mu + \epsilon_{\mu\nu} x_\nu \tag{1.23}$$

where $\epsilon_{\mu\nu} = -\epsilon_{\nu\mu}$. Under this transformation the fields that may have nonzero spin will transform as

$$\phi_k(x) \to \left(\delta_{kl} - \frac{1}{2}\epsilon_{\mu\nu}\Sigma_{kl}^{\mu\nu}\right)\phi_l(x') \tag{1.24}$$

Here $\Sigma_{kl}^{\mu\nu}$ is the spin transformation matrix, which is zero for a scalar field. The factor $\frac{1}{2}$ simplifies the final form of the spin angular momentum density.

The variation in $\mathscr{L}$ is, for this case,

$$\delta\mathscr{L} = \epsilon_{\mu\nu} x_\nu \partial_\mu \mathscr{L} \tag{1.25}$$

$$= \partial_\mu(\epsilon_{\mu\nu} x_\nu \mathscr{L}) \tag{1.26}$$

since $\epsilon_{\mu\nu}\partial_\mu x_\nu = \epsilon_{\mu\nu}\delta_{\mu\nu} = 0$ by antisymmetry.

We know, however, from an earlier result that

$$\delta\mathscr{L} = \partial_\mu\left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}\delta\phi_k\right] \tag{1.27}$$

$$= \partial_\mu\left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}\left(\epsilon_{\lambda\nu} x_\nu \partial_\lambda\phi_k - \frac{1}{2}\Sigma_{kl}^{\lambda\nu}\epsilon_{\lambda\nu}\phi_l\right)\right] \tag{1.28}$$

It follows by subtracting the two expressions for $\delta\mathscr{L}$ that if we define

$$\mathscr{M}^{\lambda\nu\nu} = \left(x^\nu g^{\lambda\mu} - x^\mu g^{\lambda\nu}\right)\mathscr{L} + \frac{\partial}{\partial(\partial_\lambda\phi_k)}\left[\left(x^\mu\partial_\nu - x^\nu\partial_\mu\right)\phi_k + \Sigma_{kl}^{\mu\nu}\phi_l\right] \tag{1.29}$$

$$= x^\mu T^{\lambda\nu} - x^\nu T^{\lambda\mu} + \frac{\partial\mathscr{L}}{\partial(\partial_\lambda\pi_k)}\Sigma_{kl}^{\mu\nu}\phi_l \tag{1.30}$$

then

$$\partial_\lambda\mathscr{M}^{\lambda\mu\nu} = 0 \tag{1.31}$$

The Lorentz generator densities may be identified as

$$\mathscr{M}^{\mu\nu} = \mathscr{M}^{0\mu\nu} \tag{1.32}$$

Their space integrals are

$$M^{\mu\nu} = \int d^3x \mathscr{M}^{\mu\nu} \tag{1.33}$$

$$= \int d^3x \left(x^\mu \mathscr{P}^\nu - x^\nu \mathscr{P}^\mu + \pi_k \Sigma_{kl}^{\mu\nu} \phi_l\right) \tag{1.34}$$

and satisfy

$$\frac{\partial}{\partial t} M^{\mu\nu} = 0 \tag{1.35}$$

The components $M^{ij}$ $(i, j = 1, 2, 3)$ are the generators of rotations and yield conservation of angular momentum. It can be seen from the expression above that the contribution from orbital angular momentum adds to a spin angular momentum part involving $\Sigma_{kl}^{\mu\nu}$.

The components $M^{0i}$ generate boosts, and the associated conservation law [3] tells us that for a field confined within a finite region of space, the "average" or center of mass coordinate moves with the uniform velocity appropriate to the result of the boost transformation (see, in particular, Hill [4]). This then completes the construction of the 10 Poincaré group generators from the Lagrangian density by use of Noether's theorem.

Now we may consider internal symmetries, that is, symmetries that are not related to space–time transformations. The first topic is global gauge invariance; in Section 1.3 we consider the generalization to local gauge invariance.

The simplest example is perhaps provided by electric charge conservation. Let the finite gauge transformation be

$$\phi_k(x) \to \phi'_k(x) = e^{-iq_k} \phi_k(x) \tag{1.36}$$

where $q_k$ is the electric charge associated with the field $\phi_k(x)$. Then every term in the Lagrangian density will contain a certain number $m$ of terms

$$\phi_{k_1}(x)\phi_{k_2}(x) \cdots \phi_{k_m}(x) \tag{1.37}$$

which is such that

$$\sum_{i=1}^{m} q_{k_i} = 0 \tag{1.38}$$

and hence is invariant under the gauge transformation. Thus the invariance implies that the Lagrangian is electrically neutral and all interactions conserve electric charge. The symmetry group is that of unitary transformations in one dimension, U(1). Quantum electrodynamics possesses this invariance: The uncharged photon has $q_k = 0$, while the electron field and its conjugate transform, respectively, according to

$$\psi \to e^{-iq\theta}\psi \tag{1.39}$$

$$\bar{\psi} \to e^{+iq\theta}\bar{\psi} \tag{1.40}$$

where $q$ is the electronic charge.

The infinitesimal form of a global gauge transformation is

$$\phi_k(x) \to \phi_k(x) - i\epsilon^i \lambda^i_{kl}\phi_l(x) \tag{1.41}$$

where we have allowed a nontrivial matrix group generated by $\lambda^i_{kl}$. Applying Noether's theorem, one then observes that

$$\delta\mathscr{L} = \partial_\mu\left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}\delta\phi_k\right] \tag{1.42}$$

$$= -i\epsilon^i\partial_\mu\left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}\lambda^i_{kl}\phi_l\right] \tag{1.43}$$

The currents conserved are therefore

$$J^i_\mu = -i\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}\lambda^i_{kl}\phi_l \tag{1.44}$$

and the charges conserved are

$$Q^i = \int d^3x j^i_0 \tag{1.45}$$

$$= -i\int d^3x\pi_k\lambda^i_{kl}\phi_l \tag{1.46}$$

satisfying

$$\frac{\partial}{\partial t}Q^i = 0 \tag{1.47}$$

The global gauge group has infinitesimal generators $Q_i$; in the simplest case, as in quantum electrodynamics, where the gauge group is U(1), there is only one such generator $Q$ of which the electric charges $q_k$ are the eigenvalues.

## 1.3
## Local Gauge Invariance

In common usage, the term *gauge field theory* refers to a field theory that possesses a local gauge invariance. The simplest example is provided by quantum electrodynamics, where the Lagrangian is

$$\mathscr{L} = \bar{\psi}(i\partial\!\!\!/ - eA\!\!\!/ - m)\psi - \frac{1}{4}F_{\mu\nu}F_{\mu\nu} \tag{1.48}$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\nu \tag{1.49}$$

Here the slash notation denotes contraction with a Dirac gamma matrix: $A\!\!\!/ \equiv \gamma_\mu A_\mu$. The Lagrangian may also be written

$$\mathscr{L} = \bar{\psi}(iD\!\!\!\!/ - m)\psi - \frac{1}{4}F_{\mu\nu}F_{\mu\nu} \tag{1.50}$$

where $D_\mu\psi$ is the covariant derivative (this terminology will be explained shortly)

$$D_\mu\psi = \partial_\mu\psi + ieA_\mu\psi \tag{1.51}$$

The global gauge invariance of quantum electrodynamics follows from the fact that $\mathscr{L}$ is invariant under the replacement

$$\psi \to \psi' = e^{i\theta}\psi \tag{1.52}$$

$$\bar{\psi} \to \bar{\psi}' = e^{-i\theta}\bar{\psi} \tag{1.53}$$

where $\theta$ is a constant; this implies electric charge conservation. Note that the photon field, being electrically neutral, remains unchanged here.

The crucial point is that the Lagrangian $\mathscr{L}$ is invariant under a much larger group of local gauge transformations, given by

$$\psi \to \psi' = e^{i\theta(x)}\psi \tag{1.54}$$

$$\bar{\psi} \to \bar{\psi}' = e^{-i\theta(x)}\bar{\psi} \tag{1.55}$$

$$A_\mu \to A'_\mu = A_\mu - \frac{1}{e}\partial_\mu\theta(x) \tag{1.56}$$

Here the gauge function $\theta(x)$ is an arbitrary function of $x$. Under the transformation, $F_{\mu\nu}$ is invariant, and it is easy to check that

$$\bar{\psi}'(i\partial\!\!\!/ - eA\!\!\!/')\psi' = \bar{\psi}(i\partial\!\!\!/ - eA\!\!\!/)\psi \tag{1.57}$$

so that $\bar{\psi}\not{D}\psi$ is invariant also. Note that the presence of the photon field is essential since the derivative is invariant only because of the compensating transformation of $A_\mu$. By contrast, in global transformations where $\theta$ is constant, the derivative terms are not problematic.

Note that the introduction of a photon mass term $-m^2 A_\mu A_\mu$ into the Lagrangian would lead to a violation of local gauge invariance; in this sense we may say that physically the local gauge invariance corresponds to the fact that the photon is precisely massless.

It is important to realize, however, that the requirement of local gauge invariance does not imply the existence of the spin-1 photon, since we may equally well introduce a derivative

$$A_\mu = \partial_\mu \Lambda \tag{1.58}$$

where the scalar $\Lambda$ transforms according to

$$\Lambda \to \Lambda' = \Lambda - \frac{1}{e}\theta \tag{1.59}$$

Thus to arrive at the correct $\mathscr{L}$ for quantum electrodynamics, an additional assumption, such as renormalizability, is necessary.

The local gauge group in quantum electrodynamics is a trivial Abelian U(1) group. In a classic paper, Yang and Mills [5] demonstrated how to construct a field theory locally invariant under a non-Abelian gauge group, and that is our next topic.

Let the transformation of the fields $\phi_k(x)$ be given by

$$\delta\phi_k(x) = -i\theta^i(x)\lambda^i_{kl}\phi_l(x) \tag{1.60}$$

so that

$$\phi_k(x) \to \phi'_k(x) = \Omega_{kl}\phi_l \tag{1.61}$$

with

$$\Omega_{kl} = \delta_{kl} - i\theta^i(x)\lambda^i_{kl} \tag{1.62}$$

where the constant matrices $\lambda^i_{kl}$ satisfy a Lie algebra ($i, j, k = 1, 2, \ldots, n$)

$$\left[\lambda^i, \lambda^j\right] = i c_{ijk}\lambda^k \tag{1.63}$$

and where the $\theta^i(x)$ are arbitrary functions of $x$.

Since $\Omega$ depends on $x$, a derivative transforms as

$$\partial_\mu\phi_k \to \Omega_{kl}(\partial_\mu\phi_l) + (\partial_\mu\Omega_{kl})\phi_l \tag{1.64}$$

We now wish to construct a *covariant* derivative $D_\mu\phi_k$ that transforms according to

$$D_\mu\phi_k \to \Omega_{kl}(D_\mu\phi_l) \tag{1.65}$$

To this end we introduce $n$ gauge fields $A^i_\mu$ and write

$$D_\mu \phi_k = (\partial_\mu - igA_\mu)\phi_k \tag{1.66}$$

where

$$A_\mu = A^i_\mu \lambda^i \tag{1.67}$$

The required transformation property follows provided that

$$(\partial_\mu \Omega)\phi - igA'_\mu \Omega\phi = -ig(\Omega A_\mu)\phi \tag{1.68}$$

Thus the gauge field must transform according to

$$A_\mu \to A'_\mu = \Omega A_\mu \Omega^{-1} - \frac{i}{g}(\partial_\mu \Omega)\Omega^{-1} \tag{1.69}$$

Before discussing the kinetic term for $A^i_\mu$ it is useful to find explicitly the infinitesimal transformation. Using

$$\Omega_{kl} = \delta_{kl} - i\lambda^i_{kl}\theta^i \tag{1.70}$$

$$\Omega^{-1}_{kl} = \delta_{kl} + i\lambda^i_{kl}\theta^i \tag{1.71}$$

one finds that

$$\lambda^i_{kl} A'^i_\mu = \Omega_{km}\lambda^i_{mn} A^i_\mu \left(\Omega^{-1}\right) - \frac{i}{g}(\partial_\mu \Omega_{km})\left(\Omega^{-1}\right)_{ml} \tag{1.72}$$

so that (for small $\theta^i$)

$$\lambda^i_{kl}\delta A^i_\mu = i\theta^j\left[\lambda^i, \lambda^j\right]_{kl} A^i_\mu - \frac{1}{g}\lambda^i_{kl}\partial_\mu\theta^i \tag{1.73}$$

$$= -\frac{1}{g}\lambda^i_{kl}\partial_\mu\theta^i - c_{ijm}\theta^j A^i_\mu \lambda^m_{kl} \tag{1.74}$$

This implies that

$$\delta A^i_\mu = -\frac{1}{g}\partial_\mu\theta^i + c_{ijk}\theta^j A^k_\mu \tag{1.75}$$

For the kinetic term in $A^i_\mu$ it is inappropriate to take simply the four-dimensional curl since

$$\delta\left(\partial_\mu A^i_\mu - \partial_\nu A^i_\mu\right) = c_{ijk}\theta^i\left(\partial_\mu A^k_\nu - \partial_\nu A^k_\mu\right) + c_{ijk}\left[\left(\partial_\mu\theta^j\right)A^k_\nu - \left(\partial_\nu\theta^j\right)A^k_\mu\right] \tag{1.76}$$

whereas the transformation property required is

$$\delta F^i_{\mu\nu} = c_{ijk}\theta^j F^k_{\mu\nu} \tag{1.77}$$

Thus $F^i_{\mu\nu}$ must contain an additional piece and the appropriate choice turns out to be

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g c_{ijk} A^j_\mu A^k_\nu \tag{1.78}$$

To confirm this choice, one needs to evaluate

$$g c_{ijk}\delta\left(A^j_\mu A^k_\nu\right) = -c_{ijk}\left[\left(\partial_\mu\theta^j\right)A^k_\nu - \left(\partial_\nu\theta^j\right)A^k_\mu\right]$$

$$+ g\left(c_{ijk}c_{jlm}\theta^l A^m_\mu A^k_\nu + c_{ijk}A^j_\mu c_{klm}\theta^l A^m_\nu\right) \tag{1.79}$$

The term in parentheses on the right-hand side may be simplified by noting that an $n \times n$ matrix representation of the gauge algebra is provided, in terms of the structure constants, by

$$\left(\lambda^i\right)_{jk} = -i c_{ijk} \tag{1.80}$$

Using this, we may rewrite the last term as

$$g A^m_\mu A^n_\nu \theta^j (c_{ipn}c_{pjm} + c_{imp}c_{pjn}) = g A^m_\mu A^n_\nu \theta^j \left[\lambda^i, \lambda^j\right]_{mn} \tag{1.81}$$

$$= i g A^m_\mu A^n_\nu \theta^j c_{ijk}\lambda^k_{mn} \tag{1.82}$$

$$= g A^m_\mu A^n_\nu \theta^j c_{ijk}c_{kmn} \tag{1.83}$$

Collecting these results, we deduce that

$$\delta F^i_{\mu\nu} = \delta\left(\partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g c_{ijk} A^j_\mu A^k_\nu\right) \tag{1.84}$$

$$= c_{ijk}\theta^j\left(\partial_\mu A^k_\nu - \partial_\nu A^k_\mu + g c_{klm} A^l_\mu A^m_\nu\right) \tag{1.85}$$

$$= c_{ijk}\theta^j F^k_{\mu\nu} \tag{1.86}$$

as required. From this it follows that

$$\delta\left(F^i_{\mu\nu}F^i_{\mu\nu}\right) = 2c_{ijk}F^i_{\mu\nu}\theta^j F^k_{\mu\nu} \tag{1.87}$$

$$= 0 \tag{1.88}$$

so we may use $-\frac{1}{4}F^i_{\mu\nu}F^i_{\mu\nu}$ as the kinetic term.

To summarize these results for construction of a Yang–Mills Lagrangian: Start with a globally gauge-invariant Lagrangian

$$\mathscr{L}(\phi_k, \partial_\mu \phi_k) \tag{1.89}$$

then introduce $A^i_\mu$ ($i = 1, \dots, n$, where the gauge group has $n$ generators). Define

$$D_\mu \phi_k = \left(\partial_\mu - igA^i_\mu \lambda^i\right)\phi_k \tag{1.90}$$

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + gc_{ijk} A^j_\mu A^k_\nu \tag{1.91}$$

The transformation properties are ($A_\mu = A^i_\mu \lambda^i$)

$$\phi' = \Omega\phi \tag{1.92}$$

$$A'_\mu = \Omega A_\mu \Omega^{-1} - \frac{i}{g}(\partial_\mu \Omega)\Omega^{-1} \tag{1.93}$$

The required Lagrangian is

$$\mathscr{L}(\phi_k, D_\mu \phi_k) - \frac{1}{4} F^i_{\mu\nu} F^i_{\mu\nu} \tag{1.94}$$

When the gauge group is a direct product of two or more subgroups, a different coupling constant $g$ may be associated with each subgroup. For example, in the simplest renormalizable model for weak interactions, the Weinberg–Salam model, the gauge group is $SU(2) \times U(1)$ and there are two independent coupling constants, as discussed later.

Before proceeding further, we give a more systematic derivation of the locally gauge invariant $\mathscr{L}$, following the analysis of Utiyama [6] (see also Glashow and Gell-Mann [7]). In what follows we shall, first, deduce the forms of $D_\mu \phi_k$ and $F^i_{\mu\nu}$ (merely written down above), and second, establish a formalism that could be extended beyond quantum electrodynamics and Yang–Mills theory to general relativity.

The questions to consider are, given a Lagrangian

$$\mathscr{L}(\phi_k, \partial_\mu \phi_k) \tag{1.95}$$

invariant globally under a group $G$ with $n$ independent constant parameters $\theta^i$, then, to extend the invariance to a group $G'$ dependent on local parameters $\theta^i(x)$:

1. What new (gauge) fields $A^p(x)$ must be introduced?
2. How does $A^p(x)$ transform under $G'$?
3. What is the form of the interaction?
4. What is the new Lagrangian?

We are given the global invariance under

$$\delta\phi_k = -iT^i_{kl}\theta^i\phi_l \quad (1.96)$$

with $i = 1, 2, \ldots, n$ and $T^i$ satisfying

$$\left[T^i, T^j\right] = ic_{ijk}T^k \quad (1.97)$$

where

$$c_{ijk} = -c_{jik} \quad (1.98)$$

and

$$c_{ijl}c_{lkm} + c_{jkl}c_{lim} + c_{kil}c_{ljm} = 0 \quad (1.99)$$

Using Noether's theorem, one finds the $n$ conserved currents

$$J^i_\mu = \frac{\partial\mathscr{L}}{\partial\phi_k}T^i_{kl}\partial_\mu\phi_l \quad (1.100)$$

$$\partial_\mu J^i_\mu = 0 \quad (1.101)$$

These conservation laws provide a necessary and sufficient condition for the invariance of $\mathscr{L}$.

Now consider

$$\delta\phi_k = -iT^i_{kl}\theta^i(x)\phi_l(x) \quad (1.102)$$

This local transformation does not leave $< J$ invariant:

$$\delta\mathscr{L} = -i\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}T^i_{kl}\phi_l\partial_\mu\partial_i \quad (1.103)$$

$$\neq 0 \quad (1.104)$$

Hence it is necessary to add new fields $A'^p$ $(p = 1, \ldots, M)$ in the Lagrangian, which we write as

$$\mathscr{L}(\phi_k, \partial_\mu\phi_k) \to \mathscr{L}'\left(\phi_k, \partial_\mu\phi_k, A'^p\right) \quad (1.105)$$

Let the transformation of $A'^p$ be

$$\delta A'^p = U^i_{pq}\theta^i A'^q + \frac{1}{g}C^{jp}_\mu\partial_\mu\phi^j \quad (1.106)$$

Then the requirement is

$$\delta\mathscr{L} = \left[-i\frac{\partial\mathscr{L}'}{\partial\phi_k}T^j_{kl}\phi_l - i\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}T^j_{kl}\partial_\mu\phi_l + \frac{\partial\mathscr{L}'}{\partial A'^p}U^j_{pq}A'^p\right]\theta^j$$

$$+\left[-i\frac{\partial\mathscr{L}'}{\partial(\partial_\mu\phi_k)}T^j_{kl}\phi_l + \frac{1}{g}\frac{\partial\mathscr{L}'}{\partial A'^p}C^{pj}_\mu\right]\partial_\mu\theta^j \tag{1.107}$$

$$= 0 \tag{1.108}$$

Since $\theta^j$ and $\partial_\mu\theta^j$ are independent, the coefficients must vanish separately. For the coefficient of $\partial_\mu\theta^i$, this gives $4n$ equations involving $A'^p$ and hence to determine the $A'$ dependence uniquely, one needs $4n$ components. Further, the matrix $C^{pj}_\mu$ must be nonsingular and possess an inverse

$$C^{jp}_\mu C^{-1jq}_\mu = \delta_{pq} \tag{1.109}$$

$$C^{jp}_\mu C^{-1j'p}_\mu = g_{\mu\nu}\delta_{jj'} \tag{1.110}$$

Now we define

$$A^j_\mu = -gC^{-1jp}_\mu A'^p \tag{1.111}$$

Then

$$i\frac{\partial\mathscr{L}'}{\partial(\partial_\mu\phi_k)}T^i_{kl}\phi_l + \frac{\partial\mathscr{L}'}{\partial A^i_\mu} = 0 \tag{1.112}$$

so only the combination

$$D_\mu\phi_k = \partial_\mu\phi_k - iT^i_{kl}\phi_l A^i_\mu \tag{1.113}$$

occurs in the Lagrangian

$$\mathscr{L}'\big(\phi_k, \partial_\mu\phi_k, A'^p\big) = \mathscr{L}''(\phi_k, D_\mu\phi_k) \tag{1.114}$$

It follows from this equality of $\mathscr{L}'$ and $\mathscr{L}''$ that

$$\left.\frac{\partial\mathscr{L}''}{\partial\phi_k}\right|_{D_\mu\phi} - i\left.\frac{\partial\mathscr{L}''}{\partial(D_\mu\phi_l)}\right|_\phi T^i_{kl}A^i_\mu = \frac{\partial\mathscr{L}'}{\partial\phi_k} \tag{1.115}$$

$$\left.\frac{\partial\mathscr{L}''}{\partial(D_\mu\phi_k)}\right|_\phi = \frac{\partial\mathscr{L}'}{\partial(D_\mu\phi_k)} \tag{1.116}$$

$$ig\left.\frac{\partial\mathscr{L}''}{\partial(D_\mu\phi_k)}\right|_\phi T^a_{kl}\phi_l C^{-1ap}_\mu = \frac{\partial\mathscr{L}'}{\partial A'^p} \tag{1.117}$$

These relations may be substituted into the vanishing coefficient of $\theta^j$ occurring in $\delta\mathscr{L}'$ (above). The result is

$$0 = -i\left[\frac{\partial \mathscr{L}''}{\partial \phi_k}\bigg|_{D_\mu \phi} T^i_{kl}\phi_l + \frac{\partial \mathscr{L}''}{\partial (D_\mu \phi_k)}\bigg|_{\phi} T^i_{kl} D_\mu \phi_l\right]$$

$$+ i\frac{\partial \mathscr{L}''}{\partial \phi_k}\bigg|_{\phi}\left(\phi_l A^a_\nu \left\{i\left[T^a, T^i\right]_{kl}\lambda_{\mu\nu} + S^{ba,j}_{\mu\nu}\right\}\right) = 0 \tag{1.118}$$

where

$$S^{ba,j}_{\mu\nu} = C^{-1ap}_\mu U^j_{pq} C^{bq}_\nu \tag{1.119}$$

is defined such that

$$\delta A^a_\mu = g\delta\left(-C^{-1ap}_\mu A'^p\right) \tag{1.120}$$

$$= S^{ba,j}_{\mu\nu} A^b_\nu \theta^j - \frac{1}{g}\partial_\mu \theta^a \tag{1.121}$$

Now the term in the first set of brackets in Eq. (1.118) vanishes if we make the identification

$$\mathscr{L}''(\phi_k, D_\mu \phi_k) = \mathscr{L}(\phi_k, D_\mu \phi_k) \tag{1.122}$$

The vanishing of the final term in parentheses in Eq. (1.118) then enables us to identify

$$S^{ba,j}_{\mu\nu} = -c_{ajb} g_{\mu\nu} \tag{1.123}$$

It follows that

$$\delta A^a_\mu = c_{abc}\theta^b A^c_\mu - \frac{1}{g}\partial_\mu \theta^a \tag{1.124}$$

From the transformations $\delta A^A_\mu$ and $\delta \phi_k$, one can show that

$$\delta(D_\mu \phi_k) = \delta\left(\partial_\mu \phi_k - iT^a_{kl} A^a_\mu \phi_l\right) \tag{1.125}$$

$$= -iT^i_{kl}\theta^i (D_\mu \phi_l) \tag{1.126}$$

This shows that $D_\mu \phi_k$ transforms covariantly.

Let the Lagrangian density for the free $A^a_\mu$ field be

$$\mathscr{L}_0\left(A^a_\mu, \partial_\nu A^a_\mu\right) \tag{1.127}$$

Using

$$\delta A^a_\mu = c_{abc}\theta^b A^c_\mu - \frac{1}{g}\partial_\mu \theta^a \tag{1.128}$$

one finds (from $\delta\mathscr{L} = 0$)

$$\frac{\partial \mathscr{L}_0}{\partial A^a_\mu} c_{abc} A^c_\mu + \frac{\partial \mathscr{L}_0}{\partial(\partial_\nu A^a_\mu)} c_{abc} \partial_\nu A^c_\mu = 0 \tag{1.129}$$

$$-\frac{\partial \mathscr{L}_0}{\partial A^a_\mu} + \frac{\partial \mathscr{L}_0}{\partial(\partial_\mu A^b_\nu)} c_{abc} A^c_\nu = 0 \tag{1.130}$$

$$\frac{\partial \mathscr{L}_0}{\partial(\partial_\nu A^a_\mu)} + \frac{\partial \mathscr{L}_0}{\partial(\partial_\mu A^a_\mu)} + \frac{\partial \mathscr{L}_0}{\partial(\partial_\mu A^a_\nu)} = 0 \tag{1.131}$$

From the last of these three it follows that $\partial_\mu A^a_\mu$ occurs only in the antisymmetric combination

$$A^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu \tag{1.132}$$

Using the preceding equation then gives

$$\frac{\partial \mathscr{L}_0}{\partial A^a_\mu} = \frac{\partial \mathscr{L}_0}{\partial A^b_{\mu\nu}} c_{abc} A^c_\nu \tag{1.133}$$

so the only combination occurring is

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g c_{abc} A^b_\mu A^c_\nu \tag{1.134}$$

Thus, we may put

$$\mathscr{L}_0\big(A^a_\mu, \partial_\nu A^a_\mu\big) = \mathscr{L}'_0\big(A^a_\mu, F^a_{\mu\nu}\big) \tag{1.135}$$

Then

$$\left.\frac{\partial \mathscr{L}_0}{\partial(\partial_\nu A^a_\mu)}\right|_A = \left.\frac{\partial \mathscr{L}'_0}{\partial F^a_{\mu\nu}}\right|_A \tag{1.136}$$

$$\left.\frac{\partial \mathscr{L}_0}{\partial A^a_\mu}\right|_{\partial_\mu A} = \left.\frac{\partial \mathscr{L}'_0}{\partial A^a_\mu}\right|_F + \left.\frac{\partial \mathscr{L}'_0}{\partial(\partial F^b_{\mu\nu})}\right|_A c_{abc} A^c_\nu \tag{1.137}$$

But one already knows that

$$\left.\frac{\partial \mathscr{L}_0}{\partial A^a_\mu}\right|_{\partial_\mu A} = \frac{\partial \mathscr{L}'_0}{\partial F^b_{\mu\nu}} c_{abc} A^c_\nu \tag{1.138}$$

and it follows that $\mathscr{L}'_0$ does not depend explicitly on $A^a_\mu$.

$$\mathscr{L}_0(A_\mu, \partial_\nu A_\mu) = \mathscr{L}''_0\big(F^a_{\mu\nu}\big) \tag{1.139}$$

Bearing in mind both the analogy with quantum electrodynamics and renormalizability we write

$$\mathscr{L}_0''(F^a_{\mu\nu}) = -\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} \tag{1.140}$$

When all structure constants vanish, this then reduces to the usual Abelian case. The final Lagrangian is therefore

$$\mathscr{L}(\phi_k, D_\mu\phi_k) - \frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} \tag{1.141}$$

Defining matrices $M^i$ in the adjoint representation by

$$M^i_{ab} = -ic_{iab} \tag{1.142}$$

the transformation properties are

$$\delta\phi_k = -iT^i_{kl}\theta^i\phi_l \tag{1.143}$$

$$\delta A^a_\mu = -iM^i_{ab}\theta^i A^b_\mu - \frac{1}{g}\partial_\mu\theta^a \tag{1.144}$$

$$\delta(D_\mu\phi_k) = -iT_{kl}\theta^i(D_\mu\phi_l) \tag{1.145}$$

$$\delta F^a_{\mu\nu} = -M^i_{ab}\theta^i F^b_{\mu\nu} \tag{1.146}$$

Clearly, the Yang–Mills theory is most elegant when the matter fields are in the adjoint representation like the gauge fields because then the transformation properties of $\phi_k$, $D_\mu\phi_k$ and $F^a_{\mu\nu}$ all coincide. But in theories of physical interest for strong and weak interactions, the matter fields will often, instead, be put into the fundamental representation of the gauge group.

Let us give briefly three examples, the first Abelian and the next two non-Abelian.

**Example 1 (Quantum Electrodynamics).** For free fermions

$$\mathscr{L}\bar{\psi}(i\partial\!\!\!/ - m)\psi \tag{1.147}$$

the covariant derivative is

$$D_\mu\psi = \partial_\mu\psi + ieA_\mu\psi \tag{1.148}$$

This leads to

$$\mathscr{L}(\psi, D_\mu\psi) - \frac{1}{4}F_{\mu\nu}F_{\mu\nu} = \bar{\psi}(i\partial\!\!\!/ - eA\!\!\!/ - m)\psi - \frac{1}{4}F_{\mu\nu}F_{\mu\nu} \tag{1.149}$$

**Example 2 (Scalar $\phi^4$ Theory with $\phi^a$ in Adjoint Representation).** The globally invariant Lagrangian is

$$\mathscr{L} = \frac{1}{2}\partial_\mu\phi^a\partial - \mu\phi^a - \frac{1}{2}\mu^2\phi^a\phi^a - \frac{1}{4}\lambda(\phi^a\phi^a)^2 \tag{1.150}$$

One introduces

$$D_\mu \phi^a = \partial_\mu \phi^a - g c_{abc} A^b \mu \phi^c \quad (1.151)$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g c_{ac} A^b_\mu A^c_\nu \quad (1.152)$$

and the appropriate Yang–Mills Lagrangian is then

$$\mathscr{L} = \frac{1}{2}\left(D_\mu \phi^a\right)\left(D_\mu \phi^a\right) - \frac{1}{4} F^a_{\mu\nu} F^a_{\mu\nu} - \frac{1}{2}\mu^a \phi^a \phi^a - \frac{1}{4}\left(\phi^a \phi^a\right)^2 \quad (1.153)$$

**Example 3 (Quantum Chromodynamics).** Here the quarks $\psi_k$ are in the fundamental (three-dimensional) representation of SU(3). The Lagrangian for free quarks is

$$\mathscr{L} \bar{\psi}_k (i\partial\!\!\!/ - m)\psi_k \quad (1.154)$$

We now introduce

$$D_\mu \psi_k = \partial_\mu \psi_k - \frac{1}{2} g \lambda^i_{kl} A^i_\mu \psi_l \quad (1.155)$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f_{abc} A^b_\mu A^c_\nu \quad (1.156)$$

and the appropriate Yang–Mills Lagrangian is

$$\mathscr{L} \bar{\psi}(i D\!\!\!\!/ - m)\psi - \frac{1}{4} F^a_{\mu\nu} F^a \mu\nu \quad (1.157)$$

If a flavor group (which is *not* gauged) is introduced, the quarks carry an additional label $\psi^a_k$, and the mass term becomes a diagonal matrix $m \to -M_a \delta_{ab}$.

The advantage of this Utiyama procedure is that it may be generalized to include general relativity (see Utiyama [6], Kibble [8], and more recent works [9–12]).

Finally, note that any mass term of the form $+m_i^2 A^i_\mu A^i_\mu$ will violate the local gauge invariance of the Lagrangian density $\mathscr{L}$. From what we have stated so far, the theory must contain $n$ massless vector particles, where $n$ is the number of generators of the gauge group; at least, this is true as long as the local gauge symmetry is unbroken.

## 1.4 Nambu–Goldstone Conjecture

We have seen that the imposition of a non-Abelian local gauge invariance appears to require the existence of a number of massless gauge vector bosons equal to the number of generators of the gauge group; this follows from the fact that a mass term $+\frac{1}{2}^2 A^i_\mu A^i_\mu$ in $\mathscr{L}$ breaks the local invariance. Since in nature only one massless spin-1 particle—the photon—is known, it follows that if we are to exploit a local gauge group less trivial than U(1), the symmetry must be broken.

Let us therefore recall two distinct ways in which a symmetry may be broken. If there is exact symmetry, this means that under the transformations of the group the Lagrangian is invariant:

$$\delta\mathscr{L} = 0 \tag{1.158}$$

Further, the vacuum is left invariant under the action of the group generators (charges) $Q_i$:

$$Q_i|0\rangle = 0 \tag{1.159}$$

From this, it follows that all the $Q_i$ commute with the Hamiltonian

$$\left[Q^i, H\right] = 0 \tag{1.160}$$

and that the particle multiplets must be mass degenerate.

The first mechanism to be considered is explicit symmetry breaking, where one adds to the symmetric Lagrangian ($\mathscr{L}_0$) a piece ($\mathscr{L}_1$) that is noninvariant under the full symmetry group $G$, although $\mathscr{L}_1$ may be invariant under some subgroup $G'$ of $G$. Then

$$\mathscr{L} = \mathscr{L}_0 = c\mathscr{L}_1 \tag{1.161}$$

and under the group transformation,

$$\delta\mathscr{L}_0 = 0 \tag{1.162}$$

$$\delta\mathscr{L}_1 \neq 0 \tag{1.163}$$

while

$$Q_i|0\rangle \to 0 \quad \text{as } c \to 0 \tag{1.164}$$

The explicit breaking is used traditionally for the breaking of flavor groups SU(3) and SU(4) in hadron physics.

The second mechanism is spontaneous symmetry breaking (perhaps more appropriately called *hidden symmetry*). In this case the Lagrangian is symmetric,

$$\delta\mathscr{L} = 0 \tag{1.165}$$

but the vacuum is not:

$$Q_i|0\rangle \neq 0 \tag{1.166}$$

This is because as a consequence of the dynamics the vacuum state is degenerate, and the choice of one as the physical vacuum breaks the symmetry. This leads to nondegenerate particle multiplets.

It is possible that both explicit and spontaneous symmetry breaking be present. One then has

$$\mathscr{L} = \mathscr{L}_0 + c\mathscr{L}_1 \tag{1.167}$$

$$\delta L = 0 \tag{1.168}$$

$$\delta L_1 \neq 0 \tag{1.169}$$

but

$$Q^i|0\rangle \neq 0 \quad \text{as } c \to 0 \tag{1.170}$$

An example that illustrates all of these possibilities is the infinite ferromagnet, where the symmetry in question is rotational invariance. In the paramagnetic phase at temperature $T > T_c$ there is exact symmetry; in the ferromagnetic phase, $T < T_c$, there is spontaneous symmetry breaking. When an external magnetic field is applied, this gives explicit symmetry breaking for both $T > T_c$ and $T < T_c$.

Here we are concerned with Nambu and Goldstone's well-known conjecture [13–15] that when there is spontaneous breaking of a continuous symmetry in a quantum field theory, there must exist massless spin-0 particles. If this conjecture were always correct, the situation would be hopeless. Fortunately, although the Nambu–Goldstone conjecture applies to global symmetries as considered here, the conjecture fails for local gauge theories because of the Higgs mechanism described in Section 1.5.

It is worth remarking that in the presence of spontaneous breakdown of symmetry the usual argument of Noether's theorem that leads to a conserved charge breaks down. Suppose that the global symmetry is

$$\phi_k \to \phi_k - iT^i_{kl}\phi_l\theta^i \tag{1.171}$$

Then

$$\partial_\mu j^i_\mu = 0 \tag{1.172}$$

$$J^i_\mu = -i\left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}T^i_{kl}\phi_l\right] \tag{1.173}$$

but the corresponding *charge*,

$$Q^i = \int d^3x j^i_0 \tag{1.174}$$

will not exist because the current does not fall off sufficiently fast with distance to make the integral convergent.

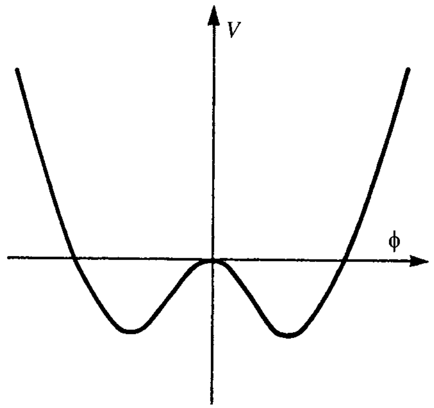



**Figure 1.1** Potential function $V(\phi)$.

The simplest model field theory [14] to exhibit spontaneous symmetry breaking is the one with Lagrangian

$$\mathscr{L} = \frac{1}{2}\left(\partial_\mu\phi\partial_\mu\phi - m_0^2\phi^2\right) - \frac{\lambda_0}{24}\phi^4 \tag{1.175}$$

For $m_0^2 > 0$, one can apply the usual quantization procedures, but for $m_0^2 < 0$, the potential function

$$V(\phi) = \frac{1}{2}m_0^2\phi^2 + \frac{\lambda_0}{24}\phi^4 \tag{1.176}$$

has the shape depicted in Fig. 1.1. The ground state occurs where $V'(\phi_a) = 0$, corresponding to

$$\phi_0 = \pm\chi = \pm\sqrt{\frac{-6m_0^2}{\lambda_0}} \tag{1.177}$$

Taking the positive root, it is necessary to define a shifted field $\phi'$ by

$$\phi = \phi' + \chi \tag{1.178}$$

Inserting this into the Lagrangian $\mathscr{L}$ leads to

$$\mathscr{L} = \frac{1}{2}\left(\partial_\mu\phi'\partial_\mu\phi' + 2m_0^2\phi'^2\right) - \frac{1}{6}\lambda_0\chi\phi'^3 - \frac{\lambda_0}{24}\phi'^4 + \frac{3m_0^4}{\lambda_0} \tag{1.179}$$

The $(\text{mass})^2$ of the $\phi'$ field is seen to be $-2m_0^2 < 0$, and this Lagrangian may now be treated by canonical methods. The symmetry $\phi \to -\phi$ of the original Lagrangian has disappeared. We may choose either of the vacuum states $\phi = \pm\chi$ as the physical vacuum without affecting the theory, but once a choice of vacuum is made, the reflection symmetry is lost. Note that the Fock spaces built on the two possi-

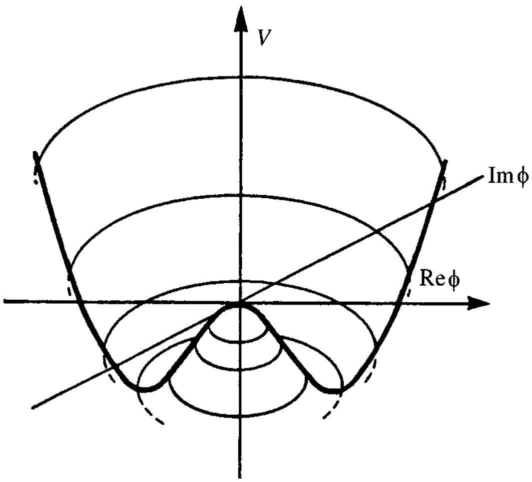



**Figure 1.2** Potential function $V(\phi, \phi^*)$.

ble vacua are mutually orthogonal so that the artificial restoration of the original symmetry by superimposing the different vacua is not meaningful.

More interesting is the generalization to a continuous U(1) symmetry in the Lagrangian [14]

$$\mathscr{L} = \partial_\mu \phi^* \partial_\mu \phi - m_0^2 \phi^* \phi - \frac{\lambda_0}{6}(\phi^* \phi)^2 \tag{1.180}$$

where

$$\phi = \frac{\phi_1 + i\phi_2}{\sqrt{2}} \tag{1.181}$$

The Lagrangian is invariant under the global symmetry

$$\phi \to e^{i\theta}\phi \tag{1.182}$$

The potential function

$$V(\phi, \phi^*) = m_0^2 \phi^* \phi + \frac{\lambda_0}{6}(\phi^* \phi)^2 \tag{1.183}$$

has the characteristic shape indicated in Fig. 1.2. The origin is an unstable point and there are an infinite number of degenerate vacua where $V' = 0$, corresponding to

$$|\phi|^2 = -\frac{3m_0^2}{\lambda_0} = |\chi|^2 \tag{1.184}$$

Let us choose the phase of $\chi$ to be real and shift fields according to

$$\phi_1 = \phi_1' + \sqrt{2}\chi \tag{1.185}$$

$$\phi_2 = \phi_2' \tag{1.186}$$

Then, in terms of $\phi'$, the Lagrangian becomes

$$\begin{aligned}\mathcal{L} &= \frac{1}{2}\left(\partial_\mu\phi_1'\partial_\mu\phi_1' + 2m_0^2\phi_1'^2\right) + \frac{1}{2}\partial_\mu\phi_2'\partial_\mu\phi_2' \\ &\quad - \sqrt{2}\frac{\lambda_0\chi}{6}\phi_1'\left(\phi_1'^2 + \phi_2'^2\right) - \frac{\lambda_0}{24}\left(\phi_1'^2 + \phi_2'^2\right)^2 + \frac{3}{2}\frac{m_0^4}{\lambda_0}\end{aligned} \tag{1.187}$$

so that the field $\phi_1$ has $(\text{mass})^2 = -2m_0^2 > 0$ and the field $\phi_2'$ has zero mass. This is an example of a Nambu–Goldstone boson associated with spontaneous breaking of a continuous symmetry.

Intuitively, one can understand the situation as follows. In the vacuum all the state vectors are lined up with the same phase and magnitude $|\phi| = \chi$. Oscillations are then of two types: One is in magnitude, giving rise to the massive quanta of type $\phi_1'$; one is in phase. When all the $\phi(x)$ rotate by a common phase, however, there is no change in energy; this is precisely the origin of the Goldstone mode represented in this example by $\phi_2'$.

Before going to the general case, let us consider the more complicated example of $\mathrm{O}(n)$ symmetry for the Lagrangian

$$\mathcal{L}\frac{1}{2}\partial_\mu\phi_k\partial_\mu\phi_k - \frac{1}{2}\mu^2\phi_k\phi_k - \frac{\lambda}{4}(\phi_k\phi_k)^2 \tag{1.188}$$

with $k = 1, 2, \ldots, n$ and $\phi_k$ real. The corresponding potential function

$$V(\phi) = \frac{1}{2}\mu^2\phi_k\phi_k + \frac{\lambda}{4}(\phi_k\phi_k)^2 \tag{1.189}$$

has a ring of minima where

$$\phi^2 = v^2 = -\frac{\mu^2}{\lambda} \tag{1.190}$$

Let us *choose* the physical vacuum such that

$$\langle 0|\phi_k|0\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ v \end{pmatrix} \tag{1.191}$$

Now the little group of this vector is clearly $\mathrm{O}(n-1)$. Let us write the $\frac{1}{2}n(n-1)$ generators of $0(n)$ as

$$(L_{ij})_{kl} = -i(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}) \tag{1.192}$$

satisfying

$$[L_{ij}, L_{kl}] = (\delta_{ik}L_{jl} + \delta_{jl}L_{ik} - \delta_{jk}L_{il} - \delta_{il}L_{jk}) \tag{1.193}$$

Of these there are $(n-1)$ given by

$$k_i L_{in} \tag{1.194}$$

which do *not* leave $\langle\phi\rangle$ invariant. We then reparametrize the $n$-component field $\phi_k$ as

$$\phi = \exp\left(i\sum_{i=1}^{n-1}\xi_i\frac{k_i}{v}\right)\begin{pmatrix}0\\ \vdots\\ 0\\ v+\eta\end{pmatrix} \tag{1.195}$$

in terms of the $(n-1)$ fields $\xi_i$, and the field $\eta$. The action of $k_i$ on the vector $v_i = v\delta_{in}$ is given by

$$(k_i v)_j = v(L_{in})_{jl}\delta_{in} \tag{1.196}$$

$$= -iv\delta_{ij} \tag{1.197}$$

Thus to lowest order, one has

$$\phi = \begin{pmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_{n-1}\\ v+\eta\end{pmatrix} \tag{1.198}$$

Inserting this into the Lagrangian, one finds that

$$\mathscr{L} = \frac{1}{2}(\partial_\mu\eta\partial_\mu\eta + \partial_\mu\xi_i\partial_\mu\xi_i) - \frac{1}{2}\mu^2(v+\eta)^2$$

$$- \frac{1}{4}\lambda(v+\eta)^4 + \text{higher-order terms} \tag{1.199}$$

Thus the $\eta$ has $(\text{mass})^2$ given by

$$\mu^2 + 3v^2\lambda = -2\mu^2 > 0 \tag{1.200}$$

and the $(n-1)$ fields $\xi_i$ are massless Goldstone bosons. Note that the number of Nambu–Goldstone bosons is equal to the difference in the number of generators of the original symmetry $O(n)$ and the final symmetry $O(n-1)$:

$$\frac{1}{2}n(n-1) - \frac{1}{2}(n-1)(n-2) = (n-1) \tag{1.201}$$

This is an example of a general theorem that we now prove.

Consider the general Lagrangian

$$\mathscr{L}(\phi_k, \partial_\mu \phi_k) = \frac{1}{2}\partial_\mu \phi_k \partial_\mu \phi_k - V(\phi_k) \tag{1.202}$$

and let the global invariance be under the transformations

$$\delta\phi_k = -iT^i_{kl}\theta^i\phi_l \tag{1.203}$$

$$\delta\mathscr{L} = 0 \tag{1.204}$$

Given that the potential is separately invariant, then

$$0 = \delta V \tag{1.205}$$

$$= \frac{\partial V}{\partial \phi_k}\delta\phi_k \tag{1.206}$$

$$= -i\frac{\partial V}{\partial \phi_k}T^i_{kl}\theta^i\phi_l \tag{1.207}$$

$$\frac{\partial V}{\partial \phi_k}T^i_{kl}\phi_l = 0 \tag{1.208}$$

Differentiation gives

$$\frac{\partial^2 V}{\partial \phi_k \partial \phi_m}T^i_{kl}\phi_l + \frac{\partial V}{\partial \phi_k}T^i_{km} = 0 \tag{1.209}$$

At a minimum of $V$, suppose that $\phi_k = v_k$; then

$$\left.\frac{\partial^2 V}{\partial \phi_k \partial \phi_m}\right|_{\phi_k = v_k} T^i_{kl}v_l = 0 \tag{1.210}$$

Expanding around the minimum

$$V(\phi_k) = V(v_k) + \frac{1}{2}\frac{\partial^2 V}{\partial \phi_l \partial \phi_m}(\phi_l - v_l)(\phi_m - v_m) + \text{higher-order terms} \tag{1.211}$$

Thus the mass matrix is

$$M^2_{mk} = M^2_{km} = \frac{\partial^2 V}{\partial \phi_k \partial \phi_m} \tag{1.212}$$

Therefore,

$$\left(M^2\right)_{mk}T^i_{kl}v_l = 0 \quad \text{all } i \tag{1.213}$$

Now suppose that there exists a subgroup $G'$ of $G$ with $n'$ generators, leaving the vacuum invariant

$$T^i_{kl} v_l = 0, \quad i = 1, 2, \ldots, n' \tag{1.214}$$

For the remaining $(n - n')$ generators of $G$,

$$T^i_{kl} v_l \neq 0 \tag{1.215}$$

Choosing the $v_l$ to be linearly independent, it follows that $M^2$ has $(n - n')$ Nambu–Goldstone bosons. Thus, in general, the number of Nambu–Goldstone bosons is equal to the number of broken generators.

If one restricts one's attention to tree diagrams, there may be exceptions to the counting rule above. These occur if the potential has a higher symmetry than the Lagrangian and can be illustrated by an elementary example. Let the Lagrangian be

$$\begin{aligned}\mathscr{L} = &\frac{1}{2}\partial_\mu \phi_k \partial_\mu \phi_k - \frac{1}{2} m_0^2 \phi_k \phi_k - \frac{\lambda_0}{8}(\phi_k \phi_k)^2 \\ &+ \bar{\psi}_a (i\partial\!\!\!/ - m)\psi_a + g\epsilon_{abc}\phi^a \bar{\psi}_b \psi_c\end{aligned} \tag{1.216}$$

where $k = 1, 2, \ldots, 8$ and $a, b, c = 1, 2, 3$. The invariance is clearly $O(3) \times O(5)$. The potential function

$$V(\phi) = \frac{1}{2} m_0^2 \phi_k \phi_k + \frac{\lambda_0}{8}(\phi_k \phi_k)^2 \tag{1.217}$$

has a minimum at

$$\langle\phi\rangle = \sqrt{-\frac{2m_0^2}{\lambda_0}} \tag{1.218}$$

Putting $\phi_k = \delta_{k8}\langle\phi\rangle$, we define

$$\sigma = \phi_8 - \langle\phi\rangle \tag{1.219}$$

$$\pi^k = \phi^k, \quad k = 1, 2, \ldots, 7 \tag{1.220}$$

Then

$$V = \frac{\lambda_0}{8}\left(\pi^k \pi^k + \sigma^2 + 2\sigma\langle\phi\rangle\right)^2 - \frac{m_0^4}{2\lambda_0} \tag{1.221}$$

Hence there are seven massless fields $\pi^k$ despite the fact that the number of broken generators in breaking $O(3) \times O(5)$ to $O(3) \times O(4)$ is only four. When loop corrections are made, the three fields $\pi_1$, $\pi_2$ and $\pi_3$ will develop a mass, and the general theorem stated earlier will become valid. The fields $\pi_1$, $\pi_2$, and $\pi_3$, which are "accidentally" massless in the tree approximation, are called *pseudo-Nambu–Goldstone bosons*.

So far we have simply given numerous examples of the Nambu–Goldstone phenomenon. Now we give two general proofs that have been constructed; these are instructive because when we turn to gauge theories, one will be able to identify precisely which assumptions involved in these proofs are violated. The first proof is due to Gilbert [16] (this work was a rebuttal to an earlier "dis-proof" by Klein and Lee [17]).

Let the generator of an internal symmetry be

$$Q = \int d^3x j_0 \tag{1.222}$$

and let this transform one scalar $\phi(x)$ into another $\phi_2(x)$, both being formed from the field operators of the theory. Thus

$$i \int d^3x \left[ j_0(x), \phi_1(x) \right] = \phi_2(y) \tag{1.223}$$

Suppose that the symmetry is spontaneously broken because

$${}_0\langle \phi_2(x) \rangle_0 = \text{constant} \neq 0 \tag{1.224}$$

Thus

$$i \int d^3x_0 \, {}_0\langle \left[ j_0(x), \phi_1(0) \right] \rangle_0 \neq 0 \tag{1.225}$$

Now Lorentz invariance dictates that the most general representation of the related Fourier transform is

$$i \int d^3x e_0^{ik\cdot x} \langle \left[ j_\mu(x), \phi_1(0) \right] \rangle_0 = \epsilon(k_0) k_\mu \rho_1(k^2) + k_\mu \rho_2(k^2) \tag{1.226}$$

Here $k_\mu$ represents the four-momentum of intermediate states inserted into the commutator. Using current conservation then leads to

$$\left[ \epsilon(k_0)\rho_1(k^2) + \rho_2(k^2) \right] k^2 = 0 \quad \text{for all } k_\mu \tag{1.227}$$

It follows that $\rho_1$ and $\rho_2$ vanish for all $k^2 \neq 0$ and the most general solution is

$$\rho_1 = C_1 \delta(k^2) \tag{1.228}$$

$$\rho_2 = C_2 \delta(k^2) \tag{1.229}$$

Now integrate over $k_0$ on both sides and use

$$\delta(k_0^2 - \mathbf{k}^2) = \frac{1}{2|\mathbf{k}|} \left[ \delta(k_0 - |\mathbf{k}|) + \delta(k_0 + |\mathbf{k}|) \right] \tag{1.230}$$

Letting $\mathbf{k} \to 0$, one obtains

$$2\pi \langle \phi_2 \rangle = \frac{1}{2}\left[(C_1 + C_2) - (-C_1 + C_2)\right] \tag{1.231}$$

$$= C_1 \tag{1.232}$$

so that

$$C_1 \neq 0 \tag{1.233}$$

and hence there must exist massless spin-0 intermediate state. Note that manifest Lorentz invariance is a crucial assumption in this proof.

A second proof, due to Jona-Lasinio [18], uses a quite different approach. Take a general Lagrangian

$$\mathscr{L}_0(\phi_k, \partial_\mu \phi_k) \tag{1.234}$$

and add to it a coupling to external sources:

$$\mathscr{L}_1 = J_k \phi_k \tag{1.235}$$

Now construct the generating functional of all Green's functions of the theory:

$$W[J_k] = {}_0\!\left\langle T\left(\exp\left\{i \int d^4x [\mathscr{L}_0 + \mathscr{L}_1]\right\}\right)\right\rangle_0 \tag{1.236}$$

The connected Green's functions are generated by

$$Z[J_k] = -i \ln W[J_k] \tag{1.237}$$

The Legendre transform of $Z[J_k]$ is

$$\Gamma[\phi_k] = Z[J_k] - \int d^4x\, J_k \phi_k \tag{1.238}$$

and generates one-particle-irreducible (1PI) graphs. In particular, the inverse proper two-particle Green's function is given by

$$\Delta^{-1}_{mn}(x - y) = \frac{\delta^2 \Gamma}{\delta \phi_m(x) \delta \phi_n(y)} \tag{1.239}$$

This is easily demonstrated as follows:

$$\frac{\delta Z}{\delta J_k(x)} = \langle \Phi_k(x) \rangle = \phi_k(x) \tag{1.240}$$

$$\frac{\delta^2 Z}{\delta J_k(x) \delta \phi_l(y)} = \delta_{kl} \delta(x - y) \tag{1.241}$$

$$= \int d^4\xi \frac{\delta^2 Z}{\delta J_k(x)\delta J_m(\xi)} \frac{\delta J_m(\xi)}{\delta\phi_l(y)} \tag{1.242}$$

$$= \int d^4\xi \, \Delta_{km}(x-\xi) \frac{\delta^2 \Gamma}{\delta\phi_m(\xi)\delta\phi_l(y)} \tag{1.243}$$

where we used

$$J_m(\xi) = \frac{\delta\Gamma}{\delta\phi_m(\xi)} \tag{1.244}$$

The result for $\Delta^{-1}_{mn}(x-y)$ follows, as required.

Let there be (global) symmetry under the transformations

$$\phi_k \to \phi_k - i^i_{kl}\theta^i\phi_l \tag{1.245}$$

Then

$$\delta J_k(x) = \int d^4\xi \frac{\delta J_k(x)}{\delta\phi_l(\xi)} \delta\phi_l(\xi) \tag{1.246}$$

$$= -i\theta^i \int d^4\xi \frac{\delta^2\Gamma}{\delta\phi_k(\xi)\delta\phi_l(\xi)} T^i_{lm}\phi_m(\xi) \tag{1.247}$$

Now for a stationary point, $\delta J_k = 0$ and hence

$$\Delta^{-1}_{kl} T^i_{lm}\phi_m = 0 \quad \text{all } i \tag{1.248}$$

Thus for all $i$ for which $T^i_{kl}\phi_m \neq 0$ there is a zero eigenvalue of $\Delta^{-1}_{kl}$ and hence a massless particle. This gives, as before, a number of Nambu–Goldstone bosons equal to the number of broken generators. (For more axiomatic proofs, see Refs. [19] and [20].)

As we shall discuss shortly, the Nambu–Goldstone conjecture is invalid in the case of gauge theories. Since there are no massless spin-0 states known, the exact application of the conjecture is physically irrelevant. But the pion has abnormally low mass and can be regarded as an approximate Nambu–Goldstone boson, as first emphasized by Nambu [13]. Although it is not a gauge theory, it is worth briefly mentioning the linear sigma model that illustrates this view of the pion and the related idea of partially conserved axial current (PCAC) [14, 21]. (Further details of the PCAC hypothesis and its consequences are provided by Refs. [22] and [23].)

The linear sigma model (Gell-Mann and Levy [21]) has Lagrangian

$$\mathscr{L} = \frac{1}{2}\partial_\mu\phi_k\partial_\mu\phi_k - \frac{1}{2}m_0^2\phi_k\phi_k - \frac{1}{4}\lambda_0(\phi_k\phi_k)^2 + c\sigma \tag{1.249}$$

where $k = 1, 2, 3, 4$ and $\sigma = \phi_4$. For $c \to 0$, $\mathscr{L}$ is O(4) invariant. For $c \neq 0$ there is explicit symmetry breaking and a corresponding partially conserved current.

$$\delta\mathscr{L} = c\delta\sigma \tag{1.250}$$

$$= -ic\theta^i T^i_{4l}\phi_l \tag{1.251}$$

$$= \partial_\mu\left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi_k)}\delta\phi_k\right] \tag{1.252}$$

so that

$$\partial_\mu\big[(\partial_\mu\phi_k)T^i_{km}\phi_m\big] = cT^i_{4l}\phi_l \tag{1.253}$$

Putting $l = 1, 2, 3$ and writing $(\phi^1, \phi^2, \phi^3) = \boldsymbol{\pi}$, one finds that

$$\partial_\mu\boldsymbol{\pi}^5_\mu = c\boldsymbol{\pi} \tag{1.254}$$

where

$$\mathbf{j}^5_\mu = \boldsymbol{\pi}(\partial_\mu\sigma) - \sigma(\partial_\mu\boldsymbol{\pi}) \tag{1.255}$$

is the axial-vector current. To deduce this result, it is easiest to use the explicit form for the o$(N)$ generators $T^i_{km}$ somewhat earlier as

$$\big(T^{ij}\big)_{kl} = -i(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}) \tag{1.256}$$

The potential function is

$$V = \frac{1}{2}m_0^2(\phi_k\phi_k) + \frac{1}{4}(\phi_k\phi_k)^2 - c\sigma \tag{1.257}$$

This has a stationary value for

$$\boldsymbol{\pi} = 0 \tag{1.258}$$

$$\sigma = u \tag{1.259}$$

where

$$\big(m_0^2 + \lambda_0 u^2\big)u = c \tag{1.260}$$

Shifting fields from $\sigma$ to $s$, where

$$\sigma = u + s \tag{1.261}$$

we find that

$$\mathscr{L} = \frac{1}{2}(\partial_\mu s\partial_\mu s + \partial_\mu\boldsymbol{\pi}\,\partial_\mu\boldsymbol{\pi}) - \frac{1}{2}m_\sigma^2 s^2 - \frac{1}{2}m_\pi^2\boldsymbol{\pi}^2 - \frac{1}{4}\lambda_0 us\big(\boldsymbol{\pi}^2 + s^2\big)$$
$$+ \left(\frac{1}{2}m_0^2u^2 + \frac{3}{4}\lambda_0 u^4\right) - \frac{1}{4}\lambda\big(\boldsymbol{\pi}^2 + s^2\big)^2 \tag{1.262}$$

where

$$m_\sigma^2 = m_0^2 + 3\lambda_0 u^2 \tag{1.263}$$

$$m_\pi^2 = m_0^2 + \lambda_0 u^2 \tag{1.264}$$

In the limit $c \to 0$, *either*

$$u = 0 \tag{1.265}$$

$$m_\sigma^2 = m_\pi^2 = m_0^2 > 0 \tag{1.266}$$

*or*

$$u \neq 0 \tag{1.267}$$

$$m_\pi^2 = m_0^2 + \lambda_0 u^2 = 0 \tag{1.268}$$

$$m_\sigma^2 = 2\lambda_0 u^2 > 0 \quad \text{for } \lambda_0 > 0 \tag{1.269}$$

When $c \neq 0$ in the second (spontaneous breaking) case,

$$m_\pi^2 = \frac{c}{u} \tag{1.270}$$

$$c = m_\pi^2 \langle\sigma\rangle \tag{1.271}$$

so that the pion is an approximate Nambu–Goldstone boson, and the axial current is partially conserved according to

$$\partial_\mu \mathbf{j}_\mu^5 = \langle\sigma\rangle m_\pi^2 \boldsymbol{\pi} \tag{1.272}$$

## 1.5
## Higgs Mechanism

If the Nambu–Goldstone conjecture were applicable to local gauge theories, the situation would be hopeless since the choice would be between unwanted massless vectors (exact symmetry) and unwanted massless scalars (spontaneous breaking).

Fortunately, gauge theories circumvent the difficulty. The suspicion that this might happen was already present from nonrelativistic examples. Although several nonrelativistic cases where the Nambu–Goldstone modes occur are well known—phonons in crystals and in liquid helium [24], and magnons in a ferromagnet [25]—it turns out that in the Bardeen–Cooper–Schrieffer (BCS) theory [26] of superconductivity, the massless excitations are absent. Because of the long-range Coulomb forces between Cooper pairs, however, there is an energy gap and no massless phonons, only massive plasmalike excitations. This phenomenon led Anderson [27] to speculate that a similar escape from Nambu–Goldstone bosons might exist for gauge theories.

The first simple model to illustrate how gauge theories evade the Nambu–Goldstone conjecture was written by Higgs [29–31]. The Goldstone model has Lagrangian ($m_0^2 < 0$)

$$\mathscr{L}(\phi, \partial_\mu \phi) = \partial_\mu \phi^* \partial_\mu \phi - m_0^2 \phi^* \phi - \frac{\lambda_0}{6}(\phi^* \phi)^2 \tag{1.273}$$

Using the results given earlier, we can make a locally gauge-invariant generalization as

$$\mathscr{L}(\phi_k, D_\mu \phi_k) - \frac{1}{4} F_{\mu\nu} F_{\mu\nu} \tag{1.274}$$

$$D_\mu \phi = \partial_\mu \phi + ie A_\mu \phi \tag{1.275}$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.276}$$

This is the Higgs model. It is invariant under

$$\phi' = e^{i\theta(x)} \phi \tag{1.277}$$

$$A'_\mu = A_\mu - \frac{1}{e} \partial_\mu \theta(x) \tag{1.278}$$

As derived earlier, the minimum of the potential occurs at

$$|\phi|^2 = -\frac{3m_0^2}{\lambda_0} = \frac{1}{2} v^2 \tag{1.279}$$

We can reparametrize $\phi$ according to

$$\phi = \exp\left(\frac{i\xi}{v}\right) \frac{v+n}{\sqrt{2}} \tag{1.280}$$

$$= \frac{1}{\sqrt{2}}(v + \eta + i\xi + \text{higher orders}) \tag{1.281}$$

Substituting this into the Lagrangian gives

$$\begin{aligned} \mathscr{L} = &-\frac{1}{4} F_{\mu\nu} F_{\mu\nu} + \frac{1}{2} \partial_\mu \eta \partial_\mu \eta \\ &+ \frac{1}{2} \partial_\mu \xi \partial_\mu \xi + m_0^2 \eta^2 + \frac{1}{2} e^2 v^2 A_\mu A_\mu + ev A_\mu \partial_\mu \xi \\ &+ \text{higher orders} \end{aligned} \tag{1.282}$$

The particle spectrum can now be made obvious by the gauge transformation with gauge function $\theta(x) = -\xi/v$:

$$A'_\mu = A_\mu + \frac{1}{ev}\partial_\mu \xi \tag{1.283}$$

$$\phi' = \frac{v+\eta}{\sqrt{2}} \tag{1.284}$$

giving

$$\begin{aligned}\mathscr{L} = &-\frac{1}{4}F_{\mu\nu}F_{\mu\nu} + \frac{1}{2}\left(\partial_\mu\eta\partial_\mu\eta + 2m_0^2\eta^2\right)\\ &+\frac{1}{2}e^2v^2A'_\mu A'_\mu - \frac{1}{6}\lambda_0 v\eta^3 - \frac{1}{24}\lambda_0\eta^4\\ &+\frac{1}{2}e^2A'^2_\mu\eta(2v+\eta) - \frac{1}{4}v^2m_0^2\end{aligned} \tag{1.285}$$

There are no massless particles! The vector field has $(\text{mass})^2 = e^2v^2$ and the $\eta$ field has $(\text{mass})^2 = -2m0^2 > 0$. The $\xi$, field, known as a *would-be Nambu–Goldstone boson*, has been gauged away to become the longitudinal mode of the massive vector $A'_\mu$. Thus the number of degrees of freedom is conserved.

Concerning the proofs of the Nambu–Goldstone conjecture given in Section 1.4, the gauge freedom means that one cannot formulate a local gauge theory so that there are manifest Lorentz invariance *and* positivity of norms (both are assumed in the proofs above).

Now we turn to a non-Abelian example where the gauge group is SU(2). The Lagrangian is

$$\begin{aligned}\mathscr{L} = &\frac{1}{2}\left(\partial_\mu\phi_k + g\epsilon^{klm}A^l_\mu\phi_m\right)\left(\partial_\mu\phi_k + g\epsilon_{klm}A^l_\mu\phi_m\right)\\ &- V(\phi_k\phi_k) - \frac{1}{4}F^k_{\mu\nu}F^k_{\mu\nu}\end{aligned} \tag{1.286}$$

where $k, l, m = 1, 2, 3$. Let the vacuum expectation value (VEV) of the field be

$$\phi_k = \delta_{k3}v \tag{1.287}$$

so that $T_1$ and $T_2$ are broken generators of SU(2) and $T_3$ is still good. In the absence of gauge invariance we would therefore expect two Nambu–Goldstone bosons; in the present case these become two would-be Nambu–Goldstone modes designated $\xi_1$ and $\xi_2$. The field may be reparametrized as

$$\phi = \exp\left[\frac{i}{v}(\xi_1T_1 + \xi_2T_2)\right]\begin{pmatrix}0\\0\\v+\eta\end{pmatrix} \tag{1.288}$$

To go to the U-gauge (the unitary gauge, where the particle spectrum, and unitarity, are manifest), one makes a gauge transformation with

$$\theta(x) = -\frac{i}{v}\left[T_1\xi_1(x) + T_2\xi_2(x)\right] \tag{1.289}$$

so that

$$\phi' = \begin{pmatrix} 0 \\ 0 \\ v+\eta \end{pmatrix} \tag{1.290}$$

In the Lagrangian, the term quadratic in $A'^{i}_{\mu}$ in the U-gauge is therefore given by

$$+\frac{g^2}{2}\epsilon^{kl3}A'^{l}_{\mu}\epsilon^{kl'3}A'^{l'}_{\mu}(v+\eta)^2 \tag{1.291}$$

which may be rewritten as

$$\frac{1}{2}M^2\left(A'^{1}_{\mu}A'^{1}_{\mu} + A'^{2}_{\mu}A'^{2}_{\mu}\right) \tag{1.292}$$

with $M = gv$. Thus there are two massive vectors and one massless vector ($A'^{3}_{\mu}$), together with one massive scalar ($\eta$). Again the would-be Nambu–Goldstone bosons $\xi_1$ and $\xi_2$ have been gauged away, and the number of degrees of freedom has been conserved.

For a general non-Abelian gauge group $G$ there are initially $n$ massless gauge vectors, before symmetry breaking, where $n$ is the number of generators for $G$. Suppose that the residual group $G'$ has $n'$ generators; this general case was analyzed by Kibble [31].

The symmetric Lagrangian is

$$\mathscr{L} = -\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} + \frac{1}{2}\left(\partial - \mu - igT^iA^i_{\mu}\right)\phi_k\left(\partial_{\mu} + igT^iA^i_{\mu}\right)\phi_k V(\phi) \tag{1.293}$$

Suppose that $V(\phi)$ is minimized by $\phi_k = v_k$. Without the gauge bosons we would expect $(n - n')$ Nambu–Goldstone particles. In the gauge invariant case we reparametrize by

$$\phi_k = \exp\left(\frac{i\sum_i \xi^i T^i}{v}\right)(v+\eta) \tag{1.294}$$

where the sum is over the $(n - n')$ broken generators $T^i$ and $\eta$ is an $n'$-dimensional vector taken orthogonal to the $(n - n')$ independent of $v_l$, satisfying

$$T^i_{kl}v_l \neq 0 \quad \text{all } i \tag{1.295}$$

After a gauge transformation with gauge function $\theta(x) = -i \sum_i \xi^i(x) T^i / v$, the quadratic term in $A'^i_\mu$ is of the form

$$-\frac{1}{2} g^2 \left(T^i v, T^j v\right) A'^i_\mu A'^j_\mu \tag{1.296}$$

After diagonalization, this leads to $(n - n')$ massive vector states. The remaining $n'$ vectors remain massless.

To summarize the Kibble counting [31], the number of would-be Nambu–Goldstone bosons is equal to the number of broken generators; these modes are absorbed to make an equal number of massive vectors. The other gauge bosons remain massless.

## 1.6 Summary

The use of Higgs scalars to induce spontaneous symmetry breaking of gauge theories has the merit that one can use perturbation theory for small couplings, and since the theories are renormalizable, one can compute systematically to arbitrary accuracy, at least in principle.

But the introduction of Higgs scalars is somewhat arbitrary and unaesthetic, especially as there is no experimental support for their existence. They may be avoided by dynamical symmetry breaking where a scalar bound state develops a nonvanishing vacuum expectation value; in this case usual perturbation theory is inapplicable and new techniques are needed.

Thus there are two optimistic possibilities:

1. Higgs scalars will be discovered experimentally; they may have so far escaped detection because they are very massive.
2. It may be shown that the Higgs scalars provide a reparametrization of the dynamical breaking mechanism. In either case, the use of the fundamental scalar fields will be vindicated.

On the other hand, in case neither of these developments occurs, it is crucial to perfect satisfactory nonperturbative methods.

## References


1 N.N. Bogoliubov and D.V. Shirkov, *Introduction to the Theory of Quantized Fields*, Wiley–Interscience, New York, 1959, p. 19.

2 J.D. Bjorken and S.D. Drell, *Relativistic Quantum Fields*, McGraw–Hill, New York, 1965, p. 17.

3 M.H.L. Pryce, *Proc. R. Soc.* A195, 62 (1948).

4 E.L. Hill, *Rev. Mod. Phys.* 23, 253 (1951).

5 C.N. Yang and R.L. Mills, *Phys. Rev.* 96, 191 (1954).

6 R. Utiyama, *Phys. Rev.* 101, 1597 (1956).

7 S.L. Glashow and M. Gell-Mann, *Ann. Phys.* 15, 437 (1961).

8 T.W.B. Kibble, *J. Math. Phys.* 2, 212 (1961).

9 C.N. Yang, *Phys. Rev. Lett.* 33, 445 (1974).

**10** W.T. Ni, *Phys. Rev. Lett.* 35, 319 (1975).
**11** A.H. Thompson, *Phys. Rev. Lett.* 35, 320 (1975).
**12** F. Mansouri and L.N. Chang, *Phys. Rev.* D13, 3192 (1976).
**13** Y. Nambu, *Phys. Rev. Lett.* 4, 380 (1960).
**14** J. Goldstone, *Nuovo Cimento* 19, 15 (1961).
**15** J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* 127, 965 (1962).
**16** W. Gilbert, *Phys. Rev. Lett.* 12, 713 (1964).
**17** A. Klein and B.W. Lee, *Phys. Rev. Lett.* 12, 266 (1964).
**18** G. Jona-Lasinio, *Nuovo Cimento* 34, 1790 (1964).
**19** R.F. Streater, *Proc. R. Soc.* A287, 510 (1965).
**20** D. Kastler, D.W. Robinson, and A. Swieca, *Commun. Math. Phys.* 2, 108 (1966).
**21** M. Gell-Mann and M. Levy, *Nuovo Cimento* 16, 705 (1960).
**22** B. Renner, *Current Algebras*, Pergamon Press, Elmsford, New York, 1968.
**23** S.L. Adler and R. Dahsen, *Current Algebras*, W.A. Benjamin, New York, 1969.
**24** N.N. Bogoliubov, *J. Phys. (U.S.S.R.)* 11, 23 (1947).
**25** W. Heisenberg, *Z. Phys.* 49, 619 (1928).
**26** J. Bardeen, L. Cooper, and J. Schrieffer, *Phys. Rev.* 106, 162 (1957).
**27** P.W. Anderson, *Phys. Rev.* 130, 439 (1963).
**28** P.W. Higgs, *Phys. Lett.* 12, 132 (1964).
**29** P.W. Higgs, *Phys. Rev. Lett.* 13, 508 (1964).
**30** P.W. Higgs, *Phys. Rev.* 145, 1156 (1966).
**31** T.W.B. Kibble, *Phys. Rev.* 155, 1554 (1967).

# 2 Quantization

## 2.1 Introduction

Here we study the question of quantizing Yang–Mills field theories. It turns out that the path integral technique is the most convenient method for doing this.

Historically, it was Feynman in 1962 who pointed out that the most naive guess for the Feynman rules conflicted with perturbative unitarity even at the one-loop level; he also suggested the use of fictitious ghost particles to remedy this in the one loop. The general prescription was justified by the elegant work of Faddeev and Popov in 1967.

We first give a detailed treatment of path integrals for nonrelativistic quantum mechanics, showing that they are equivalent to the use of the Schrödinger equation, and deriving the Rutherford cross-section formula as an example. The technique is then extended to the vacuum–vacuum amplitude of a field theory, where the number of degrees of freedom becomes infinite.

For Yang–Mills theory, we first give a canonical treatment in the Coulomb gauge. The Faddeev–Popov approach is then described, applicable to any gauge choice. This culminates in the required Feynman rules for a set of covariant gauges, including the Landau gauge and the Feynman gauge.

We then study one-loop corrections to the effective potential.

## 2.2 Path Integrals

Path integrals provide a method of quantizing a classical system; the approach was pioneered by Dirac [1] and particularly by Feynman [2, 3]. (References [1] and [2] are reprinted in Ref. [4].) It provides an alternative to conventional quantization methods and is more suitable for certain problems; in particular, of course, the reason for describing it here is that it is well suited for quantization of gauge field theories.

In the conventional approach to quantization, one begins with the classical action $S$, which is an integral over a Lagrangian density $\mathscr{L}$. By the Hamiltonian variational principle, one then derives the Lagrange equations of motion. One sets up the Hamiltonian $H$ and then proceeds with canonical quantization. In the path integral method, the time evolution of a quantum mechanical state is written directly in terms of the classical action $S$.

We first describe the technique in nonrelativistic quantum mechanics, where it is alternative, and equivalent, to the Schrödinger wave equation. The path integral method is useful particularly for finding scattering amplitudes but is, in general, ill suited for bound-state problems such as the hydrogen atom. As an illustrative example (for which path integrals work easily) we shall derive the Rutherford cross-section formula [5].[1] Only once we have become comfortable with path integrals in this way will we proceed to relativistic quantum mechanics.

The one-dimensional Schrödinger equation for a particle of mass $m$ moving in a potential $V(x)$ is

$$\left[-\frac{1}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]\psi(x,t) = i\frac{\partial}{\partial t}\psi(x,t) \tag{2.1}$$

Let the wavefunction be given at initial time $t = t_i$ by

$$\psi(x, t_i) = f(x) \tag{2.2}$$

Then we would like to find the appropriate evolution operator such that

$$\psi(x_f, t_f) = \int dx K(x_f, t_f; x, t_i) f(x) \tag{2.3}$$

Now it is obvious that

$$K(x_f, t_i; x, t_i) = \delta(x_f - x) \tag{2.4}$$

but what about general $t_f$?

We divide the time interval $(t_f - t_i)$ into $(n+1)$ small elements such that

$$t_0 = t_i \tag{2.5}$$

$$t_1 = t_i + \epsilon \tag{2.6}$$

$$\vdots$$

$$t_n = t_i + n\epsilon \tag{2.7}$$

$$t_f = t_i + (n+1)\epsilon \tag{2.8}$$

1) Our treatment follows that of M. Veltman, lectures given at International School of Elementary Particle Physics, Basko-Polje, September 1974.

The Lagrangian is

$$L = T - V \tag{2.9}$$

$$= \frac{1}{2}m\dot{x}^2 - V(x) \tag{2.10}$$

We can thus put the action into the discrete form

$$S = \int_{t_i}^{t_f} dt L \tag{2.11}$$

$$= \sum_{k=1}^{n+1} \epsilon \left[ \frac{m\dot{x}_k^2}{2} - V(x_k) \right] \tag{2.12}$$

$$= \sum_{k=1}^{n+1} \epsilon \left[ \frac{m}{2\epsilon^2}(x_k - x_{k-1})^2 - V(x_k) \right] \tag{2.13}$$

Hence given the $x_k(t_k)$, we may compute $S$; a typical path is indicated on an $x-t$ plot in Fig. 2.1.

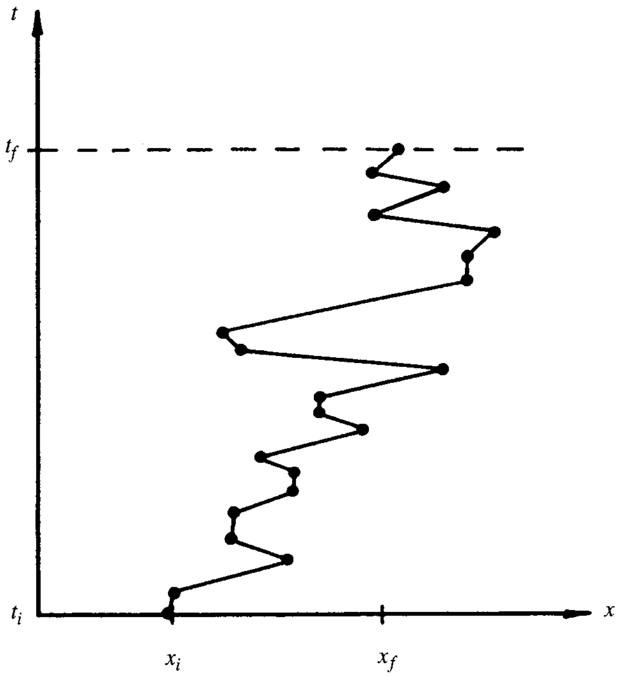



**Figure 2.1** Typical path on $x-t$ plot.

Consider now the sum over all possible paths:

$$\int_{-\infty}^{\infty} dx_1 \cdots dx_n\, e^{iS} \tag{2.14}$$

This object holds the key to the required quantum mechanical evolution operator. In the limit $\epsilon \to 0$ it will vanish in general, so it is necessary to take out a normalization factor; we thus define

$$K(x_f, t_f; x_i, t_i) = \begin{cases} \lim_{n\to\infty}\left[\frac{1}{(N(\epsilon))}\right]^{n+1} \int dx_1 \cdots dx_n\, e^{iS} & t_f > t_i \quad (2.15) \\ 0 & t_f < t_i \quad (2.16) \end{cases}$$

The appropriate value for the normalization factor, to be justified below, is

$$N(\epsilon) = \left(\frac{2\pi i\epsilon}{m}\right)^{1/2} \tag{2.17}$$

With this factor included, a convenient shorthand is

$$K(x_f, t_f; x_i, t_i) = \int Dx e^{iS} \quad t_f > t_i \tag{2.18}$$

To find the equation satisfied by $K$, consider the time $t = t_f + \epsilon$, when the position is $x_{n+2}$. Then

$$\begin{aligned} &K(x_{n+2}, t_f + \epsilon; x_i, t_i) \\ &= \frac{1}{N(\epsilon)} \int dx_{n+1} \exp\left\{i\epsilon\left[\frac{m}{2\epsilon^2}(x_{n+2} - x_{n+1})^2 - V(x_{n+2})\right]\right\} \\ &\quad \cdot K(x_{n+1}, t_f; x_i, t_i) \qquad (2.19) \\ &= \frac{1}{N(\epsilon)} \int_{-\infty}^{\infty} d\eta \exp\left[\frac{im}{2\epsilon}\eta^2 - i\epsilon V(x_{n+2})\right] \\ &\quad \cdot K(x_{n+2} + \eta, t_f; x_i, t_i) \qquad (2.20) \end{aligned}$$

where $x_{n+2} = x_{n+1} - \eta$. Expanding $K$ in a Taylor series (putting $x_{n+2} = x$ and temporarily suppressing the arguments $t_f; x_i, t_i$) yields

$$K(x+\eta) = K(x) + \eta\frac{\partial K}{\partial x}(x) + \frac{1}{2}\eta^2\frac{\partial^2 K}{\partial x^2}(x) + \cdots \tag{2.21}$$

Using the Gaussian integrals

$$\int_{-\infty}^{\infty} d\eta e^{(ia/\epsilon)\eta^2} = \sqrt{\frac{i\pi\epsilon}{a}} \tag{2.22}$$

$$\int_{-\infty}^{\infty} d\eta \, \eta e^{(ia/\epsilon)\eta^2} = 0 \tag{2.23}$$

$$\int_{-\infty}^{\infty} d\eta \, \eta^2 e^{(ia/\epsilon)\eta^2} = \frac{i\epsilon}{2a}\sqrt{\frac{i\pi\epsilon}{a}} \tag{2.24}$$

(here $a = m/2$) and expanding

$$e^{-i\epsilon V(k)} = 1 - i\epsilon V(k) + O(\epsilon^2) \tag{2.25}$$

one finds that

$$\begin{aligned} K(x, t_f + \epsilon; x_i, t_i) \\ &= \frac{1}{N(\epsilon)}\left[1 - i\epsilon V(x)\right] \\ &\quad \cdot \sqrt{\frac{2i\pi\epsilon}{m}}\left[K(x, t_f; x_i, t_i) + \frac{i\epsilon}{2m}\frac{\partial^2 K}{\partial x^2}(x, t_f; x_i, t_i)\right] + O(\epsilon^2) \end{aligned} \tag{2.26}$$

$$= K(x, t_f; x_i, t_i) + \epsilon \frac{\partial}{\partial t_f} K(x, t_f; x_i, t_i)x + O(\epsilon^2) \tag{2.27}$$

It follows that $K$ satisfies the equation

$$i\frac{\partial}{\partial t_f} K(x, t_f; x_i, t_i) = \left[-\frac{1}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right] K(x, t_f; x_i, t_i) \tag{2.28}$$

This has justified our choice of normalization $N(\epsilon)$ in Eq. (2.17). It also proves that if $\psi(x_i, t_i)$ satisfies the Schrödinger equation (2.1), so does

$$\psi(x_f, t_f) = \int dx K(x_f, t_f; x, t_i)\psi(x, t_i) \tag{2.29}$$

as required. Thus $K$ is the Green's function for the Schrödinger operator satisfying the equation

$$\left(i\frac{\partial}{\partial t_f} - V + \frac{1}{2m}\frac{\partial^2}{\partial x_f^2}\right) K(x_f, t_f; x_i, t_i) = \delta(t_f - t_i)\delta(x_f - x_i) \tag{2.30}$$

In a scattering problem, the initial wavefunction is a plane wave. In a bound-state problem, the initial wavefunction is unknown, so one has an eigenvalue problem (solving $\psi_f = \psi_i$), which is very difficult, in general.

In a problem where the potential term is small, it is most convenient to calculate $K$ in perturbation theory. One starts, in zeroth order, with the action

$$S_0 = \int_{t_i}^{t_f} dt \frac{1}{2} m\dot{x}^2 \tag{2.31}$$

whereupon

$$
\begin{aligned}
&K_0(x_f, t_f; x_i, t_i) \\
&= \int Dx e^{iS_0} &(2.32)\\
&= \lim_{n\to\infty}\left(\frac{m}{2i\pi\epsilon}\right)^{(n+1)/2} \int dx_1 \cdots dx_n \cdot \exp\left[\frac{im}{2\epsilon}\sum_{k=1}^{n+1}(x_k - x_{k-1})^2\right] &(2.33)\\
&= \lim_{n\to\infty}\sqrt{\frac{m}{2i\pi\epsilon(n+1)}}\exp\left[\frac{im}{2\epsilon(n+1)}x\right]^2 &(2.34)\\
&= \sqrt{\frac{m}{2i\pi t}}\exp\left(\frac{im}{2t}x^2\right) &(2.35)
\end{aligned}
$$

where $t = (n+1)\epsilon = t_f - t_i$.

In reaching Eq. (2.34) we have used the integral

$$
\begin{aligned}
I_n &= \int_{-\infty}^{+\infty} dx_1 \cdots dx_n \exp\{i\lambda[(x_1 - x_i)^2 + \cdots + (x_f - x_n)^2]\} &(2.36)\\
&= \sqrt{\frac{i^n\pi^n}{(n+1)\lambda^n}}\exp\left[\frac{i\lambda}{(n+1)}(x_f - x_i)^2\right] &(2.37)
\end{aligned}
$$

This is easily verified for $n = 1$ and is proved in general by induction: Assume it to be correct for $n$; then

$$
I_{n+1} = \sqrt{\frac{i^n\pi^n}{(n+1)\lambda^n}}\int_{-\infty}^{+\infty} dx_{n+1}\exp\left[\frac{i\lambda}{n+1}(x_{n+1} - x_i)^2 + i\lambda(x_f - x_{n+1})^2\right] \tag{2.38}
$$

$$
\begin{aligned}
&= \sqrt{\frac{i^n\pi^n}{(n+1)\lambda^n}}\int_{-\infty}^{+\infty} dx_{n+1}\\
&\quad\cdot\exp\left\{i\lambda\left[\frac{n+2}{n+1}y^2 - 2y(x_f - x_i) + (x_f - x_i)^2\right]\right\} &(2.39)\\
&= \sqrt{\frac{i^n\pi^n}{(n+1)\lambda^n}}\exp\left[\frac{i\lambda}{n+2}(x_f - x_i)^2\right]\int_{-\infty}^{\infty} dx \exp\left(i\lambda\frac{n+2}{n+1}z^2\right) &(2.40)\\
&= \sqrt{\frac{i^{n+1}\pi^{n+1}}{(n+2)\lambda^{n+1}}}\exp\left[i\frac{\lambda}{n+2}(x_f - x_i)^2\right] &(2.41)
\end{aligned}
$$

as required. Here we used the substitutions

$$y = x_{n+1} - x_i \tag{2.42}$$

$$z = y - \frac{n+1}{n+2}(x_f - x_i) \tag{2.43}$$

respectively.

Going into momentum space and using the formulas

$$\sqrt{\frac{\alpha}{\pi i}} \int_{-\infty}^{\infty} dp e^{ipx + i\alpha p^2} = e^{-ix^2/4\alpha} \tag{2.44}$$

$$\theta(t) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\tau \frac{e^{i\tau t}}{\tau - i\epsilon} \tag{2.45}$$

we may rewrite (with $t = t_f - t_i$, $x = x_f - x_i$)

$$K_0(x_f t_f, x_i t_i) = \frac{1}{2\pi}\theta(t) \int_{-\infty}^{\infty} dp\, e^{ipx - i(p^2/2m)t} \tag{2.46}$$

$$= \frac{1}{(2\pi)^2 i} \int_{-\infty}^{\infty} dp\, d\tau \frac{e^{ipx - i(p^2/2m)t + it\tau}}{\tau - i\epsilon} \tag{2.47}$$

$$= \frac{1}{(2\pi)^2 i} \int_{-\infty}^{\infty} dp\, dE \frac{e^{i(px - Et)}}{-E + p^2/2m - i\epsilon} \tag{2.48}$$

where $E = p^2/2m - \tau$.

Expression (2.48) represents the time development or propagation of the free wavefunction; it is thus the nonrelativistic propagator.

For the first-order correction, one includes the term linear in $V$,

$$K_1(x_f t_f; x_i t_i)$$

$$= -i \int_{t_i}^{t_f} dt \int Dx e^{iS_0} V(x,t) \tag{2.49}$$

$$= -i \left[\frac{1}{N(\epsilon)}\right]^{n+1} \sum_{l=1}^{n+1} \epsilon \int dx_1 \cdots dx_n$$

$$\cdot \exp\left[ i \sum_{k=1}^{n+1} \frac{m}{2\epsilon}(x_k - x_{k-1})^2 \right] V(x_l, t_l) \tag{2.50}$$

$$= -i \sum_{l=1}^{n+1} \epsilon \int dx_l \left[\frac{1}{N(\epsilon)}\right]^{n-l+1} \int dx_{l+1} \cdots dx_n$$

$$\cdot \exp\left[ i \sum_{k=l+1}^{n+1} \frac{m}{2\epsilon}(x_k - x_{k-1})^2 \right] V(x_l \cdot t_l) \left[ \frac{1}{N(\epsilon)} \right]^l \int dx_1 \cdots dx_{l-1}$$

$$\cdot \exp\left[ i \sum_{k=1}^{l} \frac{m}{2\epsilon}(x_k - x_{k-1})^2 \right] \tag{2.51}$$

$$= -i \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dx K_0(x_f, t_f; x, t) V(x, t) \cdot K_0(x, t; x_i, t_i) \tag{2.52}$$

The time integration can safely be extended to the full range $(-\infty, +\infty)$ because of the property that $K_0 \sim \theta(t)$.

The second and higher orders are similar; for example,

$$K_2 = (-i)^2 \int dt_1\, dt_2\, dx_1\, dx_2 K_0(x_f, t_f; x_1, t_1) V(x_1, t_1)$$

$$\cdot K_0(x_1, t_1; x_2, t_2) V(x_2, t_2) K_0(x_2, t_2; x_i, t_i) \tag{2.53}$$

In general, the $(n!)^{-1}$ of the exponential expansion is canceled by the existence of the $n!$ different time orderings. The full series reads

$$K(x_f, t_f; x_i, t_i) = K_0(x_f, t_f; x_i, t_i)$$

$$-i \sum_{n=0}^{\infty} dt\, dx K_n(x_f, t_f; x, t) V(x, t) K_0(x, t; x_i, t_i) \tag{2.54}$$

We may equivalently write a Bethe–Salpeter equation in the form

$$K(x_f, t_f; x_i, t_i) = K_0(x_f, t_f; x_i, t_i)$$

$$- i \int dt\, dx K(x_f, t_f; x, t) V(x, t) K_0(x, t; x_i, t_i) \tag{2.55}$$

The Feynman rules for this nonrelativistic single-particle case are indicated in Fig. 2.2 for configuration space. Figure 2.3 shows the diagrammatic representation of the Bethe–Salpeter equation, (2.55).

In momentum space, we define the Fourier transform of the potential as

$$V(x, t) = \int dp\, dE e^{ipx - iEt} \mathscr{V}(p, E) \tag{2.56}$$

whereupon the Feynman rules are as indicated in Fig. 2.4. Of course, the Feynman diagrams are trivial for this nonrelativistic single-particle case, but they will become very useful and nontrivial in quantum field theory.

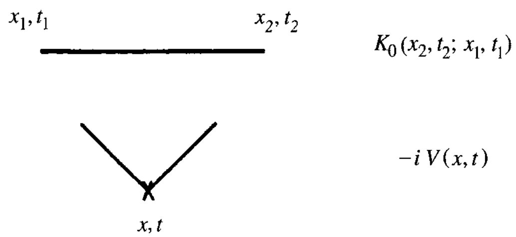



**Figure 2.2** Feynman rules for nonrelativistic particle in configuration space.


**Figure 2.3** Bethe–Salpeter equation.

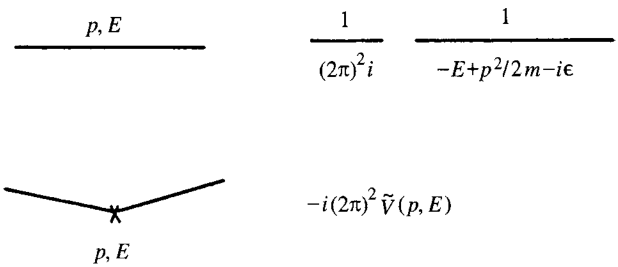



**Figure 2.4** Feynman rules for nonrelativistic particle in momentum space.

Now we may consider the calculation of an $S$-matrix element. Here, the initial and final states are plane waves that extend over all space. The $S$-matrix element is given by the overlap integral

$$\int dx_j \psi^*_{\text{out}}(x_f, t_f)\psi_{\text{in}}(x_f, t_f)$$
$$= \int dx_f\, dx_i \psi^*_{\text{out}}(x_f, t_f)K(x_f, t_f; x_i, t_i)\psi_{\text{in}}(x_i, t_i) \tag{2.57}$$

The in and out states may be normalized such that there is precisely one particle in all of space:

$$\int \psi^*_{\text{in}}\psi_{\text{in}}\, d^3x = \int \psi^*_{\text{ou}}\psi_{\text{out}}\, d^3x = 1 \tag{2.58}$$

or, in a box of volume $V$ cm$^3$, the plane-wave function is

$$\psi(x, t) = \frac{1}{\sqrt{V}}e^{i\mathbf{k}\cdot\mathbf{x}-iEt} \tag{2.59}$$

for three-momentum $\mathbf{k}$ and energy $E = k^2/2m$. The incident velocity is $k/m$ and since the density is $V^{-1}$ particles/cm$^3$, the incident flux is given by

$$\frac{k}{Vm}\text{particles/cm}^2 \cdot \text{s} \tag{2.60}$$

Let the final momentum be $\mathbf{p}$ and the $S$-matrix be $A(p, k, t)$, where $T$ is the time for which the potential is switched on so that the transition probability per second is

$$\frac{|A(p, k, T)|^2}{T} \tag{2.61}$$

The total cross section is obtained by integrating over the final-state momenta

$$\sigma_{\text{tot}} = \int d^3p \frac{V}{(2\pi)^3} \frac{Vm}{k} \frac{|A(p, k, T)|^2}{T} \tag{2.62}$$

Here the factor $V/(2\pi)^3$ is the density of plane waves in the box per unit three-momentum interval. Note that since the in and out states are plane waves, they extend over all space so that the potential must be switched on and off—very slowly to minimize energy absorption.

Thus to find, as an illustrative example, the Rutherford cross section, we use the Coulomb potential

$$V(\mathbf{x}, t) = \frac{\alpha}{r} e^{-4t^2/T^2} \tag{2.63}$$

where the fine-structure constant is $\alpha = e^2/4\pi = (137.036)^{-1}$. To find the probability per unit time we will need to divide by

$$\int_{-\infty}^{\infty} dt e^{-8t^2/T^2} = \sqrt{\frac{T^2\pi}{8}} \tag{2.64}$$

We will compute only the Born approximation (i.e., the first-order level, which uses the kernel $K_1$). The amplitude is

$$A(\mathbf{p}, \mathbf{k}, T) = \int d^3x_f\, d^3x\, dt\, d^3x_i \psi^*_{\text{out}}(x_f, t_f) K_0(\mathbf{x}_f, t_f; \mathbf{x}, t)$$

$$\cdot (-i) V(\mathbf{x}, t) K_0(\mathbf{x}, t; \mathbf{x}_i, t_i) \psi_{\text{in}}(\mathbf{x}_i, t_i) \tag{2.65}$$

Provided that $t_f \gg T$ and $t_i \ll -T$, we can drop the $\theta(t)$ and use

$$K_0 = \frac{1}{(2\pi)^3} \int d^3q \exp\left[i\mathbf{q} \cdot (\mathbf{x}_f - \mathbf{x}) - i\frac{\mathbf{q}^2}{2m}(t_f - t)\right] \tag{2.66}$$

We find that

$$A(\mathbf{p}, \mathbf{k}, T) = -i \int d^3x_f\, d^3x_i\, d^3x\, dt\, d^3q\, d^3q' \frac{1}{V} \frac{1}{(2\pi)^6} V(\mathbf{x}, t)$$

$$\cdot \exp\Big[ -i\mathbf{p} \cdot \mathbf{x}_f + i\frac{p^2}{2m} t_f + iq \cdot (\mathbf{x}_f - \mathbf{x}) - i\frac{\mathbf{q}^2}{2m}(t_f - t)$$

$$+ i\mathbf{q}' \cdot (\mathbf{x} - \mathbf{x}_i) - i\frac{\mathbf{q}'^2}{2m}(t - t_i) + i\mathbf{k} \cdot \mathbf{x}_i - i\frac{\mathbf{k}^2}{2m} t_i \Big] \tag{2.67}$$

$$= -\frac{i}{V} \int d^3x\, dt \exp\left( -i\mathbf{p} \cdot \mathbf{x} + i\frac{\mathbf{p}^2}{2m} t + i\mathbf{k} \cdot \mathbf{x} - i\frac{\mathbf{k}^2}{2m} t \right)$$

$$\cdot \frac{\alpha}{r} e^{-4t^2/T^2} \tag{2.68}$$

We need the integrals

$$\int_{-\infty}^{\infty} dt e^{iat - 4t^2/T^2} = e^{-a^2T^2/16} \int_{-\infty}^{\infty} d\tau e^{-4\tau^2/T^2} \tag{2.69}$$

$$= \sqrt{\frac{\pi T^2}{4}} e^{-a^2T^2/16} \tag{2.70}$$

with $a = (\mathbf{p}^2 - \mathbf{k}^2)/2m$ and (using a temporary cutoff at $r = M^{-1}$)

$$\int d^3x \frac{e^{i\mathbf{q}\cdot\mathbf{x}}}{r} e^{-Mr} = 2\pi \int_0^{\infty} r\, dr e^{-Mr} \int_{-1}^{+1} dz\, e^{iqrz} \tag{2.71}$$

$$= \frac{4\pi}{q^2 + M^2} \tag{2.72}$$

In Eq. (2.71) we chose the $z$-axis along $\mathbf{q} = (\mathbf{k} - \mathbf{p})$. With these integrals it follows that

$$A(\mathbf{p}, \mathbf{k}, T) = -\frac{4\pi i}{V} \frac{\alpha}{q^2 + M^2} \sqrt{\frac{\pi T^2}{4}} e^{-a^2T^2/16} \tag{2.73}$$

and hence

$$\frac{\text{probability}}{\text{time}} = |A|^2 \sqrt{\frac{8}{T^2\pi}} \tag{2.74}$$

$$= \frac{16\pi^2}{V^2} \frac{\alpha^2}{(q^2 + M^2)^2} \frac{\pi T^2}{4} e^{-a^2T^2/8} \sqrt{\frac{8}{T^2\pi}} \tag{2.75}$$

As $T$ increases we find eventually that

$$\delta(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau e^{ia\tau - 2\tau^2/T^2} \tag{2.76}$$

$$= \lim_{T\to\infty} \left( \frac{1}{2\pi} e^{-a^2T^2/8} \sqrt{\frac{\pi T^2}{2}} \right) \tag{2.77}$$

so then

$$\frac{\text{probability}}{\text{time}} = \frac{16\pi^2}{V^2} \frac{\alpha^2}{(q^2+M^2)^2} 2\pi\delta\left( \frac{\mathbf{k}^2}{2m} - \frac{\mathbf{p}^2}{2m} \right) \tag{2.78}$$

Finally, we compute the total cross section, using Eq. (2.62):

$$\sigma_{\text{tot}} = \int d^3p \frac{V}{(2\pi)^3} \frac{mV}{k} \alpha^2 \frac{32\pi^3}{V^2} \frac{1}{(q^2+M^2)^2} \cdot \delta\left( \frac{\mathbf{p}^2}{2m} - \frac{\mathbf{k}^2}{2m} \right) \tag{2.79}$$

Now, for $E = \mathbf{p}^2/2m$,

$$\int d^3p = m \int d\Omega \int \sqrt{2mE}\, dE \tag{2.80}$$

Also,

$$q^2 = 4k^2 \sin^2 \frac{\theta}{2} \tag{2.81}$$

where $\theta$ is the scattering angle. Putting all this together, we arrive at

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2 m^2}{4k^4 \sin^4(\theta/2)} \tag{2.82}$$

which is the celebrated Rutherford formula [5].

To proceed to relativistic field theory, the number of degrees of freedom must be increased from one to an infinite number. We first treat a field theory with defining Lagrangian $L(\phi)$, containing no derivatives and with no local gauge invariance, and show that we regain the same Feynman rules as from the usual canonical procedure. Gauge theories are the subject of Section 2.3.

For a single degree of freedom, we had (let us put $m = 1$; it can be restored if desired)

$$K(x_f, t_f; x_i, t_i) = \lim_{n\to\infty} \left( \frac{1}{2i\pi\epsilon} \right)^{(n+1)/2} \int_{-\infty}^{\infty} dq_1 \cdots dq_n e^{iS} \tag{2.83}$$

$$S = \sum_{k=1}^{n+1} \epsilon \left[ \frac{1}{2\epsilon^2} (q_k - q_{k-1})^2 - V(q_k) \right] \tag{2.84}$$

$$= \sum_{k=1}^{n+1} \epsilon \left[ \frac{1}{2} \dot{q}_k^2 - V(q_k) \right] \tag{2.85}$$

We may now use

$$\frac{1}{\sqrt{2i\pi\epsilon}} \exp\left( \frac{1}{2} \epsilon \dot{q}_k^2 \right) = \int \frac{dp_k}{2\pi} \exp\left[ i\epsilon \left( p_k \dot{q}_k - \frac{1}{2} p_k^2 \right) \right] \tag{2.86}$$

to rewrite

$$K = \int \frac{DpDq}{2\pi} \exp\left\{ i \int_{-\infty}^{\infty} \left[ p\dot{q} - \left( \frac{1}{2} p^2 + V \right) \right] dt \right\} \tag{2.87}$$

$$= \lim_{n\to\infty} \int \prod_{i=1}^{n} dq_i \prod_{j=1}^{n+1} \frac{dp_j}{2\pi}$$

$$\cdot \exp\left\{ i \sum_{j=1}^{n+1} \left[ p_j (q_j - q_{j-1}) - H\left( p_j, \frac{1}{2}(q_j + q_{j-1}) \right) (t_j - t_{j-1}) \right] \right\} \tag{2.88}$$

where

$$H(p,q) = \frac{1}{2} p^2 + V(q) \tag{2.89}$$

is the Hamiltonian. Thus the following are equivalent formulas for the one-degree-of-freedom case:

$$K(x_f, t_f; x_i, t_i) = \int \frac{DqDp}{2\pi} e^{i\int (p\dot{q} - H)dt} \tag{2.90}$$

$$= \int Dq e^{i\int L(q,\dot{q})dt} \tag{2.91}$$

$$= \int Dq e^{iS} \tag{2.92}$$

The generalization of Eq. (2.88) to any finite number $N$ of degrees of freedom is immediate: namely,

$$K(q_{1f}\cdots q_{Nf},t_f;q_{1f}\cdots q_{Ni},t_i)$$

$$=\int\prod_{\alpha=1}^{N}\frac{Dp_\alpha Dq_\alpha}{2\pi}\exp\left\{i\int_{t_i}^{t_f}\left[\sum_{\alpha=1}^{N}p_\alpha\dot{q}_\alpha-H(p,q)\right]dt\right\} \tag{2.93}$$

$$=\lim_{n\to\infty}\int\prod_{\alpha=1}^{N}\prod_{i=1}^{n}dq_\alpha(t_i)\prod_{i=1}^{n+1}\frac{dp_\alpha}{2\pi}(t_i)$$

$$\cdot\exp\left(i\sum_{j=1}^{n}\left\{\sum_{\alpha=1}^{N}p_\alpha(t_j)\left[q_\alpha(t_j)-q_\alpha(t_{j-1})\right]\right.\right.$$

$$\left.\left.-\epsilon H\left(p(t_j),\frac{q(t_j)+q(t_{j-1})}{2}\right)\right\}\right) \tag{2.94}$$

This is the appropriate transformation matrix (or time evolution matrix) for many degrees of freedom.

Consider now a scalar field $\phi(x)$. We subdivide space into cubes of side $\epsilon$ and volume $\epsilon^3$, labeled by the index $\alpha$ (which runs over an infinite number of values). Then define

$$\phi_\alpha=\frac{1}{\epsilon^3}\int_{V_\alpha}d^3x\phi(\mathbf{x},t) \tag{2.95}$$

The Lagrangian becomes

$$L'=\int d^3xL \tag{2.96}$$

$$=\sum_\alpha\epsilon^3L_\alpha\big(\dot{\phi}_\alpha(t),\phi_\alpha(t),\phi_{\alpha\pm\epsilon}(t)\big) \tag{2.97}$$

The momentum conjugate to $\phi_\alpha$ is

$$p_\alpha(t)=-\frac{\partial L'}{\partial\dot{\phi}_\alpha(t)}=\epsilon^3\frac{\partial L_\alpha}{\partial\dot{\phi}_\alpha(t)}=\epsilon^3\pi_\alpha(t) \tag{2.98}$$

and the Hamilton becomes:

$$H'=\sum_\alpha p_\alpha\dot{\phi}_\alpha-L' \tag{2.99}$$

$$=\sum_\alpha\epsilon^3H_\alpha \tag{2.100}$$

$$H_\alpha=\pi_\alpha\dot{\phi}_\alpha-L_\alpha \tag{2.101}$$

The transformation matrix for the field theory may now be written

$$\lim_{\substack{n\to\infty \\ \epsilon\to 0}} \int \prod_{\alpha} \prod_{i=1}^{n} d\phi_\alpha(t_i) \prod_{i=1}^{n+1} \frac{\epsilon^3}{2\pi} d\pi_\alpha(t_i)$$

$$\cdot \exp\left\{ i \sum_{j=1}^{n+1} \epsilon \sum_{\alpha} \epsilon^3 \left[ \pi_\alpha(t_j) \frac{\phi_\alpha(t_j) - \phi_\alpha(t_{j-1})}{\epsilon} \right.\right.$$

$$\left.\left. - H_\alpha\big(\pi_\alpha(t_j), \phi_\alpha(t_j), \phi_{\alpha\pm\epsilon}(t_j)\big)\right]\right\}$$

$$= \int D\phi\, D\left(\pi \frac{\epsilon^3}{2\pi}\right) \exp\left[ \int_{t_i}^{t_f} d\tau\, d^3x \left( \pi \frac{\partial \phi}{\partial \tau} - H \right) \right] \tag{2.102}$$

where in the final form we have changed the variable $t$ to $\tau = it$ to go into Euclidean space. These functional integrals are better defined in Euclidean space than in Minkowski space. It is then postulated that the Minkowski Green's functions are obtained by analytic continuation in $\tau$ of the Euclidean ones.

The generating functional may be written, with an external source $J$, as

$$W[J] = \int D\phi\, D\left(\frac{\pi\epsilon^3}{2\pi}\right)$$

$$\cdot \exp\left\{ i \int d^4x \big[ \pi(x)\dot{\phi}(x) - H(x) + J(x)\phi(x) \big] \right\} \tag{2.103}$$

Then the complete Green's functions (including the disconnected contributions) are given by

$$\left.\frac{\delta^n W[J]}{\delta J(x_1)\cdots\delta J(x_n)}\right|_{J=0} = i^n \langle 0|T\big(\phi(x_1)\cdots\phi(x_n)\big)|0\rangle \tag{2.104}$$

$$= i^n G(x_1\cdots x_n) \tag{2.105}$$

The connected Green's functions are given by

$$G_c(x_1\cdots x_n) = (-i)^{n-1} \left.\frac{\delta^n Z[J]}{\delta J(x_1)\cdots\delta J(x_n)}\right|_{J=0} \tag{2.106}$$

where

$$W[J] = \exp\big(iZ[J]\big) \tag{2.107}$$

Let us prove that $Z[J]$ generates the connected graphs only. Graphically, the general $n$-point Green's function $i^n G(x_1, \ldots, x_n)$ is composed of many connected sub-graphs. Of these, let $p_i$ be $i$-point connected graphs, such that $\sum ip_i = n$. A graphical expansion is thus given by

$$i^n G(x_1, x_2, \ldots, x_n)$$

$$= \sum_{\substack{\text{partitions } \{p_1 \cdots p_k\} \\ \sum ip_i = n}} \sum_{\substack{\text{permutations} \\ \{x_1 \cdots x_n\}}} i^{p_1} G_c(x_1) \cdots G_c(x_{p_1})$$

$$\cdots i^{p_k} G_c(x_{\sum_{i=1}^{k-1} ip_i} \cdots x_n) W[0] \tag{2.108}$$

where the $G_c$ are connected Green's functions and the factor $W[0]$ is the sum of the amplitudes of all vacuum–vacuum graphs. But from our definitions,

$$i^n G(x_1 \cdots x_n) = \frac{\delta^n}{\delta J(x_1) \cdots \delta J(x_n)} W[J] \bigg|_{J=0} \tag{2.109}$$

$$= \frac{\delta^n}{\delta J(x_1) \cdots \delta J(x_n)} e^{iZ[J]} \bigg|_{J=0} \tag{2.110}$$

This may be rewritten as

$$i^n G(x_1 \cdots x_n) = \sum_{\substack{\{p_1 \cdots p_k\} \\ \sum ip_i = n}} \sum_{\{x_1 \cdots x_n\}} \frac{\delta(iZ)}{\delta J(x_1)} \bigg|_{J=0} \cdots \frac{\delta(iZ)}{\delta J(x_{p_1})} \bigg|_{J=0}$$

$$\cdots \frac{\delta^{p_k}(iZ)}{\underset{\sum_1 ip_i + 1}{\delta J(x_{k-1}) \cdots \delta J(x_n)}} \bigg|_{J=0} e^{iZ[J]} \big|_{J=0} \tag{2.111}$$

The identification

$$i^n G_c(x_1 \cdots x_n) = \frac{\delta^n}{\delta J(x_1) \cdots \delta J(x_n)} iZ[J] \bigg|_{J=0} \tag{2.112}$$

follows by induction on $n$, from Eqs. (2.108) and (2.111). Therefore,

$$G_c(x_1 \cdots x_n) = (-i)^{n-1} \frac{\delta^n}{\delta J(x_1) \cdots \delta J(x_n)} Z[J] \bigg|_{J=0} \tag{2.113}$$

Let us now suppose that the Hamiltonian is in the form

$$H(x) = \frac{1}{2}\pi^2(x) + f(\phi, \nabla\phi) \tag{2.114}$$

Then we may perform the $\pi$ functional integral by recalling that

$$\int_{-\infty}^{\infty} \frac{dp}{2\pi} \exp\left[i\epsilon\left(p\dot{q} - \frac{1}{2}p^2\right)\right] = \frac{1}{\sqrt{2\pi i\epsilon}} e^{(1/2)i\epsilon\dot{q}^2} \tag{2.115}$$

and hence

$$\int D\left(\frac{\pi\epsilon^3}{2\pi}\right)\exp\left[i\int d^4x\left(\pi\dot{\phi}-\frac{\pi^2}{2}\right)\right]\sim\exp\left\{i\left[\int d^4x\left(\frac{1}{2}\dot{\phi}^2\right)\right]\right\} \quad (2.116)$$

Thus

$$W[J]=\int D\phi\, D\left(\frac{\pi\epsilon^3}{2\pi}\right)\exp\left[i\int d^4x(x\dot{\phi}-H+J\phi)\right] \quad (2.117)$$

$$\sim\int D\phi\exp\left[i\int d^4x(L+J\phi)\right] \quad (2.118)$$

Writing

$$L=L_0+L_I \quad (2.119)$$

$$L_0=\frac{1}{2}\left[(\partial_\mu\phi)^2-\mu^2\phi^2\right] \quad (2.120)$$

$$L_I=L_I(\phi) \quad (2.121)$$

in Euclidean space, we have the convergent integral

$$W_E[J]=\int D\phi\exp\left[-\int d^3x\,d\tau\left[\frac{1}{2}\left(\frac{\partial\phi}{\partial\tau}\right)^2 + \frac{1}{2}(\nabla\phi)^2+\mu^2\phi^2-L_I(\phi)-J\phi\right]\right] \quad (2.122)$$

To obtain the right boundary conditions, we need to add a term $\sim i\epsilon$ which ensures that positive (but not negative) frequencies are propagating forward in time. The appropriate replacement is

$$L_0\to\frac{1}{2}\left[(\partial_\mu\phi)^2-\mu^2\phi^2+i\epsilon\phi^2\right] \quad (2.123)$$

which also provides a convergence factor $\exp(-\epsilon\phi^2)$ in Minkowski space.

In zeroth order, for the free-field case, we therefore have

$$W_0[J]$$

$$=\int D\phi\exp\left\{i\int d^4x\left[\frac{1}{2}(\partial_\mu\phi)^2-\frac{1}{2}\mu^2\phi^2+\frac{1}{2}i\epsilon\phi^2+J\phi\right]\right\} \quad (2.124)$$

$$=\lim_{\epsilon\to 0}\int\prod_\alpha d\phi_\alpha\exp\left(i\sum_\alpha\epsilon^4\sum_\beta\epsilon^4\frac{1}{2}\phi_\alpha K_{\alpha\beta}\phi_\beta+\sum_\alpha\epsilon^4 J_\alpha\phi_\alpha\right) \quad (2.125)$$

where we have written a matrix form which is such that

$$\left.\begin{matrix} \alpha \to x \\ \beta \to y \end{matrix}\right\} \quad \text{as } \epsilon \to 0 \tag{2.126}$$

and

$$\lim_{\epsilon \to 0} K_{\alpha\beta} = \left(-\partial^2 - \mu^2 + i\epsilon\right)\delta^4(x-y) \tag{2.127}$$

We can perform the integral exactly in this discretized form, to obtain

$$\begin{aligned} W_0[J] = & \lim_{\epsilon \to 0} \frac{1}{\sqrt{\det K_{\alpha\beta}}} \prod_\alpha \frac{2\pi i}{\epsilon^4} \\ & \cdot \exp\left[-\frac{1}{2} i \sum_\alpha \epsilon^4 \sum_\beta \epsilon^4 J_\alpha \frac{1}{\epsilon^8} \left(K^{-1}\right)_{\alpha\beta} J_\beta\right] \end{aligned} \tag{2.128}$$

with

$$\left(K^{-1}\right)_{\alpha\beta} (K)_{\beta\gamma} = \delta_{\alpha\gamma} \tag{2.129}$$

Taking the continuum limit, which implies that

$$\frac{1}{\epsilon^4} \delta_{\alpha\beta} \to \delta(x-y) \tag{2.130}$$

$$\sum_\alpha \epsilon^4 \to \int d^4x \tag{2.131}$$

we define as the propagator

$$\frac{1}{\epsilon^8} \left(K^{-1}\right)_{\alpha\beta} \to \Delta_F(x-y) \tag{2.132}$$

Then

$$W_0[J] = \exp\left[-\frac{1}{2} i \int d^4x\, d^4y\, J(x) \Delta_F(x-y) J(y)\right] \tag{2.133}$$

where $\Delta_F$ satisfies

$$\left(-\partial^2 - \mu^2 + i\epsilon\right)\Delta_F(x-y) = \delta(x-y) \tag{2.134}$$

The solution of this is

$$\Delta_F(x-y) = \int \frac{d^4k}{(2\pi)^4} \frac{e^{-ik\cdot(x-y)}}{k^2 - \mu^2 + i\epsilon} \tag{2.135}$$

This is the Feynman propagator.

For the interacting case, we have

$$W[J] = \int D\phi \exp\left[i \int d^4x (L_0 + L_I + J\phi)\right] \tag{2.136}$$

$$= \exp\left[i \int d^4x L_I\left(-i\frac{\delta}{\delta J(x)}\right)\right]$$

$$\cdot \int \mathscr{D}\phi \exp\left[i \int d^4x (L_0 + J\phi)\right] \tag{2.137}$$

$$\sim \sum_{p=0}^{\infty} \frac{(i)^p}{p!}\left[\int d^4x L_I\left(-i\frac{\delta}{\delta J}\right)\right]^p$$

$$\cdot \exp\left[-\frac{i}{2}\int d^4x\, d^4y J(x)\Delta_F(x-y)J(y)\right] \tag{2.138}$$

The usual Wick's theorem now follows by repeated use of

$$\frac{\delta}{\delta J(x)} J(y) = \delta^4(x-y) \tag{2.139}$$

For a given number ($n$) of space-time points and given number ($p$) of vertices (i.e., $p$th order of perturbation theory), we can represent the result of computing Eq. (2.138) diagramatically. There are both connected and disconnected diagrams; examples for $n = 4$ and $p = 2$ are shown in Fig. 2.5, where the interaction is taken as $L_I = -\lambda\phi^4$. For this case, each interaction vertex involves a 4! factor because of the 4! ways the $(\delta/\delta J)^4$ can be applied. Apart from this, every possible graph occurs once in the expression. Thus, each vertex gives

$$-4! i\lambda \int d^4x \tag{2.140}$$

and each propagator gives

$$i\Delta_F(x-y) \tag{2.141}$$

Note that the $\frac{1}{2}$ factor in the $W_0[J]$ exponent goes out because of the two $J$ factors. The external lines carry propagators; to find an $S$-matrix element, we put the external lines on a mass shell, canceling the external propagators by Klein–Gordon operators.

The general rules are now clear. We write down $n$ points $x_1 \cdots x_n$, and $p$ points $x, y, z, w, v, u, \ldots$ and arrive at an expression

$$(-i\lambda)^p (4!)^p \int d^4x \cdots d^4u \sum_{\substack{\text{all} \\ \text{diagrams}}} \left[(i)^{(4p+n)/2} \text{ then } \frac{n+4p}{2} \text{ propagators}\right] \tag{2.142}$$

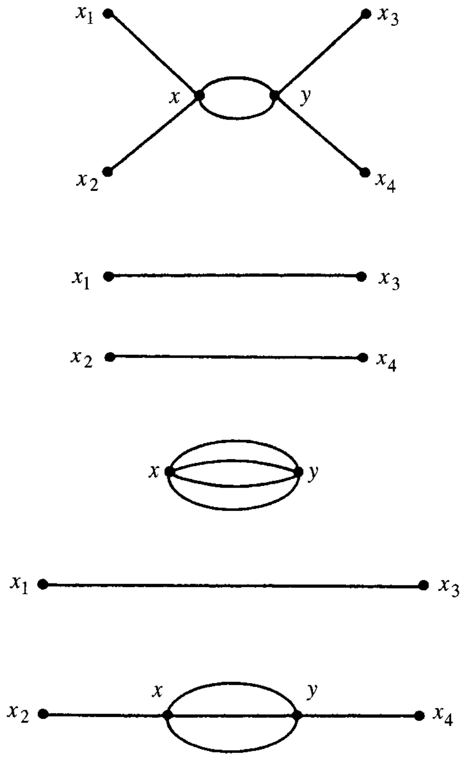



**Figure 2.5** Second-order Feynman diagrams for $\lambda\phi^4$ field theory.

The rules can be reexpressed in momentum space, but we leave that to the reader. Let us now reproduce the same Feynman rules from the canonical formalism.[2] The starting Lagrangian was

$$L = \frac{1}{2}\left[(\partial_\mu \phi)^2 - \mu^2\phi^2\right] - \lambda\phi^4 \tag{2.143}$$

The canonically conjugate momentum is

$$\pi = \frac{\partial L}{\partial \dot{\phi}} = \dot{\phi} \tag{2.144}$$

and hence the Hamiltonian is

$$H = \pi\dot{\phi} - L \tag{2.145}$$

$$= \pi^2 - \frac{1}{2}\pi^2 + \frac{1}{2}(\nabla\phi)^2 + \frac{1}{2}\mu^2\phi^2 + \lambda\phi^4 \tag{2.146}$$

2) The reader familiar with canonical quantization may skip the remainder of this section.

$$= \left[\frac{1}{2}\pi^2 + \frac{1}{2}(\nabla\phi)^2 + \frac{1}{2}\mu^2\phi^2\right] + \lambda\phi^4 \tag{2.147}$$

$$= H_0 + H_I(\phi) \tag{2.148}$$

The Heisenberg equations of motion are

$$\frac{\partial\phi}{\partial t} = i[H', \phi] \tag{2.149}$$

$$\frac{\partial\pi}{\partial t} = i[H', \pi] \tag{2.150}$$

$$H' = \int d^3x H \tag{2.151}$$

The canonical quantization of the fields proceeds by requiring that

$$\left[\phi(\mathbf{x}, t), \pi(\mathbf{x}', t)\right] = i\delta^3(x - \mathrm{x}') \tag{2.152}$$

$$\left[\phi(\mathbf{x}, t), \phi(\mathbf{x}, t)\right] = 0 \tag{2.153}$$

$$\left[\pi(\mathbf{x}, t), \pi(\mathbf{x}', t)\right] = 0 \tag{2.154}$$

Next we introduce a unitary time-evolution operator relating the in state to the interacting state:

$$\phi(x) = U^{-1}(t)\phi_{\text{in}}(x)U(t) \tag{2.155}$$

$$\pi(x) = U^{-1}(t)\pi_{\text{in}}(x)U(t) \tag{2.156}$$

Starting from

$$\phi_{\text{in}} = U\phi U^{-1} \tag{2.157}$$

and differentiating with respect to time, using $U^{-1}U = 1$ and the equations of motion, one easily finds that

$$\frac{\partial\phi_{\text{in}}}{\partial t} = \left[\dot{U}U^{-1}, \phi_{\text{in}}\right] + i\left[H(\phi_{\text{in}}), \phi_{\text{in}}\right] \tag{2.158}$$

But, by definition,

$$\frac{\partial\phi_{\text{in}}}{\partial t} = i\left[H_0(\phi_{\text{in}}), \phi_{\text{in}}\right] \tag{2.159}$$

Therefore,

$$i\dot{U}U^{-1} = H_I(t) \tag{2.160}$$

$$i\frac{\partial \dot{U}(t,t')}{\partial t} = H_I(t)U(t,t') \tag{2.161}$$

The well-known solution is [here $U(t_1, t_2) = U(t_1)U^{-1}(t_2)$]

$$U(t,t') = 1 - i\int_{t'}^{t} dt_1\, H_I(t_1)U(t_1,t') \tag{2.162}$$

$$= \sum_{n=0}^{\infty}(-i)^n \int_{t'}^{t} dt_1 \int_{t'}^{t_1} dt_2 \cdots \int_{t'}^{t_{n-1}} dt_n\, H_I(t_1)\cdots H_I(t_n) \tag{2.163}$$

$$= \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{t'}^{t} dt_1 \cdots \int_{t'}^{t_{n-1}} dt_n\, T\big(H_I(t_1)\cdots H_I(t_n)\big) \tag{2.164}$$

$$= T\exp\Big[-i\int d^4x\, H_I\big(\phi(x)\big)\Big] \tag{2.165}$$

$$= T\exp\Big[+i\int d^4x\, L_I\big(\phi(x)\big)\Big] \tag{2.166}$$

To calculate the required Green's functions, we write

$$G(x_1\cdots x_n)$$

$$= \langle 0|T\big(\phi(x_1)\cdots\phi(x_n)\big)|0\rangle \tag{2.167}$$

$$= \langle 0|T\big(U^{-1}(t_1)\phi_{\text{in}}(x_1)U(t_1,t_2)\phi_{\text{in}}(x_2)\cdots\phi_{\text{in}}(x_n)U(t_n)\big)|0\rangle \tag{2.168}$$

Using Eq. (2.166) leads us to

$$G(x_1\cdots x_n)$$

$$= \Big\langle 0\Big|T\big(\Phi_{\text{in}}(x_1)\cdots\Phi_{\text{in}}(x_n)\big)\cdot\exp\Big[i\int_{t_n}^{t1} d^4x\, L_I(\Phi_{\text{in}})\Big]\Big|0\Big\rangle \tag{2.169}$$

$$= \sum_{p=0}^{\infty}\frac{(i)^p}{p!}\int d^4x\, d^4u\ (p \text{ integrations})$$

$$\cdot\ \langle 0|T\big(\Phi_{\text{in}}(x_1)\cdots\phi_{\text{in}}(x_n)L_I\big(\phi_{\text{in}}(x)\big)\cdots L_I\big(\phi_{\text{in}}(u)\big)\big)|0\rangle \tag{2.170}$$

Now, we use Wick's algebraic theorem [6] for time-ordered products:

$$T\big(\phi_{\text{in}}(x_1)\cdots\phi_{\text{in}}(x_n)\big) =: \phi_{\text{in}}(x_1)\cdots\phi_{\text{in}}(x_n):$$

$$+\ \langle 0|T\big(\phi_{\text{in}}(x_1)\phi_{\text{in}}(x_2)\big)|0\rangle\ :\phi(x_3)\cdots\phi(x_n):$$

$$
\begin{aligned}
&+ \text{permutations} \\
&+ \cdots \\
&+ \langle 0|T\big(\phi_{\text{in}}(x_1)\phi_{\text{in}}(x_2)\big)|0\rangle \langle 0|T\big(\phi_{\text{in}}(x_3)\phi_{\text{in}}(x_4)\big)|0\rangle \\
&\cdots \langle 0|T\big(\phi_{\text{in}}(x_{n-1})\phi_{\text{in}}(x_n)\big)|0\rangle \\
&+ \text{permutations}
\end{aligned} \tag{2.171}
$$

Finally, exploiting the fact that

$$\langle 0|T\big(\phi_{\text{in}}(x_1)\phi_{\text{in}}(x_2)\big)|0\rangle = i\Delta_F(x_1 - x_2) \tag{2.172}$$

$$= i\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot(x_1-x_2)}}{k^2-\mu^2+i\epsilon} \tag{2.173}$$

we again obtain Feynman rules: For the contribution of order $p$ in $L_I$ to the $n$-point function, we draw $(n+p)$ dots, make all possible diagrams, and arrive at an expression (for $L_I = -\lambda\phi^4$)

$$(-i\lambda)^p(4!)^p \int d^4x\cdots d^4u \sum_{\substack{\text{all}\\ \text{diagrams}}} \left[(i)^{(4p+n)/2} \text{ then } \frac{n+4p}{2} \text{ propagators}\right] \tag{2.174}$$

agreeing precisely with Eq. (2.142), which was obtained from a path integral.

Thus not only have we illustrated the attractive use of the path integrals in nonrelativistic scattering problems, but we have also proved that one obtains precisely coincident results for quantizing a field theory using path integrals as one does using the conventional canonical quantization procedure.

## 2.3 Faddeev–Popov Ansatz

To obtain the correct Feynman rules and study renormalization of non-Abelian gauge field theories, it is simpler to quantize using path integrals [7–13]. The canonical quantization procedure becomes particularly awkward in this case.

Let us consider an SU(2) Yang–Mills theory [14] in the Coulomb gauge [15]. In the first-order formalism where $A_\mu^a$ and $F_{\mu\nu}^a$ are independent, the Lagrangian is

$$L = \frac{1}{4}\mathbf{F}_{\mu\nu}\cdot\mathbf{F}_{\mu\nu} - \frac{1}{2}\mathbf{F}_{\mu\nu}\cdot(\partial_\mu\mathbf{A}_\nu - \partial_\nu\mathbf{A}_\mu + g\mathbf{A}_\mu{}^\wedge\mathbf{A}_\nu) \tag{2.175}$$

with $\mathbf{a}\cdot\mathbf{b} = a^a b^a$ and $(\mathbf{a}^\wedge\mathbf{b})^a = \epsilon^{abc}a^b b^c \cdot L$ is invariant under the local gauge transformations

$$\mathbf{A}_\mu(x) \to \mathbf{A}_\mu(x) + \mathbf{U}(x)^\wedge \mathbf{A}_\mu(x) - \frac{1}{g}\partial_\mu \mathbf{U}(x) \tag{2.176}$$

$$\mathbf{F}_{\mu\nu}(x) \to \mathbf{F}_{\mu\nu}(x) + \mathbf{U}(x)^\wedge \mathbf{F}_{\mu\nu} \tag{2.177}$$

where

$$\mathbf{U}(x) = 1 - \frac{i}{2}\boldsymbol{\tau}\cdot\boldsymbol{\theta}(x) \tag{2.178}$$

The Euler–Lagrange equations of motion are

$$\frac{\partial L}{\partial \mathbf{F}_{\mu\nu}} = 0 \tag{2.179}$$

$$\frac{\partial L}{\partial \mathbf{A}_\mu} = \partial_\nu \frac{\partial L}{\partial(\partial_\nu \mathbf{A}_\mu)} \tag{2.180}$$

which give, respectively,

$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + g\mathbf{A}_\mu{}^\wedge \mathbf{A}_\nu \tag{2.181}$$

$$\partial_\mu \mathbf{F}_{\mu\nu} + g\mathbf{A}_\mu{}^\wedge \mathbf{F}_\mu \nu = 0 \tag{2.182}$$

The question now is: What are really the independent fields? These equations impose constraints and lead to possible inconsistencies (e.g., absence of conjugate momenta). It would be inconsistent to impose canonical commutators on all field components, including the dependent ones, because this would contradict the constraints, in general, at the quantum level.

Consider first the fact that

$$\frac{\partial L}{\partial(\partial_0 \mathbf{A}_\mu)} = -\mathbf{F}_{0\mu} \tag{2.183}$$

Thus, for $\mathbf{A}_i$, $\mathbf{F}_{0i}$ is the canonically conjugate momentum.

On the other hand, $L$ is independent of $\partial_0 \mathbf{A}_0$, so that $\mathbf{A}_0$ is not independent. We can immediately express $\mathbf{F}_{ij}$ from Eq. (2.181) as

$$\mathbf{F}_{ij} = \partial_i \mathbf{A}_j - \partial_j \mathbf{A}_i + g\mathbf{A}_i{}^\wedge \mathbf{A}_j \tag{2.184}$$

but there is also the relation, from Eq. (2.182),

$$(\nabla_k + g\mathbf{A}_k{}^\wedge)\mathbf{F}_{k0} = 0 \tag{2.185}$$

This implies that not all $\mathbf{F}_{k0}$, and hence not all $\mathbf{A}_k$, are independent. We may exploit local gauge invariance and impose the Coulomb gauge condition

$$\nabla_i \mathbf{A}_i = 0 \tag{2.186}$$

We separate $\mathbf{F}_{0i}$ into transverse and longitudinal parts,

$$\mathbf{F}_{0i} = \mathbf{F}_{0i}^T + \mathbf{F}_{0i}^T \tag{2.187}$$

such that $\nabla_i \mathbf{F}_{0i}^T = 0$ and

$$\nabla_i \mathbf{F}_{0i} = \nabla_i \mathbf{F}_{0i}^L \tag{2.188}$$

$$\epsilon_{ijk} \nabla_j \mathbf{F}_{0k}^L = 0 \tag{2.189}$$

Now we wish to write $\mathbf{A}_0$ and $\mathbf{F}_{0i}^L$ in terms of independent variables and hence construct the Hamiltonian as a function only of $\mathbf{F}_{0i}^T$ and $\mathbf{A}_i$.

Because of Eq. (2.189) we may write (defining $\mathbf{f}$)

$$\mathbf{F}_{0k}^L = -\nabla_k \mathbf{f} \tag{2.190}$$

Also, we define the generalized electric field

$$\mathbf{E}_i = \mathbf{F}_{0i}^T \tag{2.191}$$

Now, from Eq. (2.185) we may write

$$(\nabla_k + g\mathbf{A}_k{}^\wedge)\mathbf{F}_{k0} = D_k \mathbf{F}_{k0} \tag{2.192}$$

$$= 0 \tag{2.193}$$

Therefore,

$$D_k \nabla_k \mathbf{f} = g\mathbf{A}_i{}^\wedge \mathbf{E}_i \tag{2.194}$$

Now let us formally assume that the operator

$$O^{ab}\big(\mathbf{A}_i(x)\big) = D_k^{ab} \nabla_k \tag{2.195}$$

$$= \big[\delta^{ab} \partial_k - g\epsilon^{abc} A_k^c(x)\big] \nabla_k \tag{2.196}$$

possesses an inverse, denoted by the Green's function, satisfying

$$O^{ab}\big(A_i(x)\big) G^{bc}(x, y, A) = \delta(x - y)\delta^{ac} \tag{2.197}$$

(*Important note*: This assumption is not strictly correct, since the operator $O^{ab}$ is singular [16, 17]; however, our conclusions will be valid in perturbation theory.)

Then the solution is written

$$f^a(x, t) = g \int d^3y G^{ab}(x, y, A) \epsilon^{bcd} A_k^c(y, t) E_k^d(y, t) \tag{2.198}$$

or, symbolically,

$$\mathbf{f} = gG\mathbf{A}_k{}^{\wedge}\mathbf{E}_k \tag{2.199}$$

where $G$ denotes the integral operator.

Similarly, we may find an equation for $\mathbf{A}$:

$$\partial_0 \mathbf{A}_i = \mathbf{F}_{0i} + (\nabla_i + g\mathbf{A}_i{}^{\wedge})\mathbf{A}_0 \tag{2.200}$$

Taking the divergence, and using the Coulomb gauge condition, Eq. (2.186), results in

$$O^{ab}\big(\mathbf{A}_i(x)\big)A_0^b(x) = \nabla^2 f^a \tag{2.201}$$

so that

$$\mathbf{A}_0 = G\nabla^2\mathbf{f} \tag{2.202}$$

Since $\mathbf{A}_i$ and $\mathbf{E}_i$ are conjugate variables, the Hamiltonian density is

$$H = \mathbf{E}_i \cdot \dot{\mathbf{A}}_i - L \tag{2.203}$$

Using Eq. (2.200), we have

$$\dot{\mathbf{A}}_i = \mathbf{E}_i - \nabla_i\mathbf{f} + (\nabla_i + g\mathbf{A}_i{}^{\wedge})G\nabla^2\mathbf{f} \tag{2.204}$$

and hence

$$\int d^3x\,\mathbf{E}_i \cdot \dot{\mathbf{A}}_i = \int d^3x\big[\mathbf{E}_i^2 - \mathbf{f}\cdot\nabla^2\mathbf{f}\big] \tag{2.205}$$

$$= \int d^3x\big[\mathbf{E}_i^2 + (\nabla_i\mathbf{f})^2\big] \tag{2.206}$$

But

$$L = \frac{1}{4}\mathbf{F}_{\mu\nu}\cdot\mathbf{F}_{\mu\nu} - \frac{1}{2}\mathbf{F}_{\mu\nu}\cdot(\partial_\mu\mathbf{A}_\nu - \partial_\nu\mathbf{A}_\mu + g\mathbf{A}_\mu{}^{\wedge}\mathbf{A}_\nu) \tag{2.207}$$

$$= \frac{1}{2}(\mathbf{F}_{0k})^2 - \frac{1}{2}(\mathbf{B}_i)^2 \tag{2.208}$$

$$= \frac{1}{2}(\mathbf{E}_K - \nabla_k\mathbf{f})^2 - \frac{1}{2}(\mathbf{B}_i)^2 \tag{2.209}$$

so that, after partial integration, one obtains

$$\int d^3x L = \int d^3x\left[\frac{1}{2}\big(\mathbf{E}_i^2 - \mathbf{B}_i^2\big) + \frac{1}{2}(\nabla_k\mathbf{f})^2\right] \tag{2.210}$$

so the Hamiltonian may be written

$$H = \frac{1}{2}\int d^3x\left[\mathbf{E}_i^2 + \mathbf{B}_i^2 + (\nabla_i \mathbf{f})^2\right] \tag{2.211}$$

Here $B_i^a = \frac{1}{2}\epsilon_{ijk}F_{jk}^a$ is the generalized magnetic field, and we have used $\int d^3x(\mathbf{E}_k \cdot \nabla_k \mathbf{f}) = 0$.

The generating functional in the Coulomb gauge is thus

$$W_c[J] = \int DE_i^T DA_i^T \cdot \exp\left\{i\int d^4x\left[\mathbf{E}_k \cdot \dot{\mathbf{A}}_k - \frac{1}{2}\mathbf{E}_k^2 - \frac{1}{2}\mathbf{B}_k^2 - \frac{1}{2}(\nabla_k \mathbf{f})^2 - \mathbf{A}_k \mathbf{J}_k\right]\right\} \tag{2.212}$$

To extend the path integral to all components of $\mathbf{E}_i$ and $\mathbf{A}_i$, we make the decomposition

$$\mathbf{E}_i = \left(\delta_{ij} - \nabla_i \frac{1}{\nabla^2}\nabla_j\right)E_j + \nabla_i \frac{1}{\nabla^2}\mathbf{E}^L \tag{2.213}$$

so that

$$\mathbf{E}^L = \nabla_i \mathbf{E}_i \tag{2.214}$$

The Jacobian to pass from $\mathbf{E}_i$ to $\mathbf{E}_i^L$ and $\mathbf{E}_i^T$ is independent of the fields and is hence only an irrelevant multiplicative constant in the generating functional. Treating $\mathbf{A}_i$ similarly, one arrives at

$$W_c[J] = \int D\mathbf{E}_i\, D\mathbf{A}_i \prod_x \delta(\nabla_k \mathbf{E}_k)\delta(\nabla_k \mathbf{A}_k) \cdot \exp\left\{i\int d^4x\left[\mathbf{E}_k \cdot \dot{A}_k - \frac{1}{2}\mathbf{E}_k^2 - \frac{1}{2}\mathbf{B}_k^2 - \frac{1}{2}(\nabla_k \mathbf{f})^2 - \mathbf{A}_k \cdot \mathbf{J}_k\right]\right\} \tag{2.215}$$

The Feynman rules found in this way are noncovariant, and awkward for calculations, but with a little more effort we may arrive at a covariant form, involving path integrals over the Lorentz vector and tensor $\mathbf{A}_\mu$ and $\mathbf{F}_{\mu\nu}$ together with extra delta functions and a very important determinant (Jacobian).

Starting from Eq. (2.199) enables us to insert an integral

$$\int D\mathbf{f}\delta(\mathbf{f} - gG\mathbf{A}_k{}^\wedge\mathbf{E}_k) \tag{2.216}$$

We may also write this, using Eq. (2.194), as

$$\int D\mathbf{f}(\det M_c)\delta(\nabla_i D_i \mathbf{f} - g\mathbf{A}_k{}^\wedge\mathbf{E}_k) \tag{2.217}$$

where $M_c$ (the subscript denoting Coulomb gauge) is the matrix, dropping an irrelevant overall factor $(1/g)$,

$$M_c^{ab}(x, y) = \nabla^2\big[\delta^{ab}\delta^3(x-y) + g\epsilon^{abc}G_0(x, y)A_i^c(y)\nabla_i\big]\delta(x_0 - y_0) \tag{2.218}$$

with $G_0(x, y)$ the Abelian Coulomb Green's function satisfying

$$\nabla^2 G_0(x, y) = \delta^3(x-y) \tag{2.219}$$

We now change variables to

$$\mathbf{F}_{0i} = \mathbf{E}_i - \nabla_i \mathbf{f} \tag{2.220}$$

and write

$$D\mathbf{E}_i D\mathbf{f}\prod_x \delta(\nabla_i \mathbf{E}_i)\delta(\nabla_i D_i \mathbf{f} - g\mathbf{A}_k{}^\wedge \mathbf{E}_k)$$

$$= D\mathbf{F}_{0i} D\mathbf{f}\prod_x \delta\big(\nabla_i \mathbf{F}_{0i} + \nabla^2\mathbf{f}\big)\delta\big(\nabla^2\mathbf{f} - a\mathbf{A}_i{}^\wedge \mathbf{F}_{0i}\big) \tag{2.221}$$

$$= D\mathbf{F}_{0i} D\mathbf{f}\prod_x \delta(\nabla_i \mathbf{F}_{0i} + g\mathbf{A}_i{}^\wedge \mathbf{F}_{0i})\delta\big(\nabla^2\mathbf{f} - g\mathbf{A}_i \hat{\mathbf{F}}_{0i}\big) \tag{2.222}$$

The path integral over **f** may now be performed, using the second delta function, with Jacobian $\det(\nabla^2)$ providing only a field-independent (infinite) multiplicative constant.

Proceeding further, we rewrite

$$\prod_x \delta(\nabla_i \mathbf{F}_{0i} + g\mathbf{A}_i{}^\wedge \mathbf{F}_{0i})$$

$$= \prod_x \int \frac{d\mathbf{A}_0}{2\pi}\exp\big[i\mathbf{A}_0 \cdot (\nabla_i \mathbf{F}_{0i} + g\mathbf{A}_i{}^\wedge \mathbf{F}_{0i})\big] \tag{2.223}$$

$$\propto \int D\mathbf{A}_0 \exp\left[i\int d^4x\, \mathbf{F}_{0i}(g\mathbf{A}_0{}^\wedge \mathbf{A}_i - \nabla_i \mathbf{A}_0)\right] \tag{2.224}$$

Also,

$$\int D\mathbf{F}_{ij}\exp\left\{i\int d^4x\left[\frac{1}{4}\mathbf{F}_{ij}\cdot\mathbf{F}_{ij} - \frac{1}{2}\mathbf{F}_{ij}\cdot(\nabla_i \mathbf{A}_j - \nabla_j \mathbf{A}_i + g\mathbf{A}_i{}^\wedge \mathbf{A}_J)\right]\right\}$$

$$\propto \exp\left[-\frac{1}{4}i\int d^4x(\nabla_i \mathbf{A}_j - \nabla_j \mathbf{A}_i + g\mathbf{A}_i{}^\wedge \mathbf{A}_j)^2\right] \tag{2.225}$$

Collecting results then gives the required covariant form for the generating functional, namely

$$W_c[\mathbf{J}] = \int D\mathbf{A}_\mu \, D\mathbf{F}_{\mu\nu} (\det M_c) \prod_x \delta(\nabla_i \mathbf{A}_i)$$

$$\cdot \exp\left[ i \int d^4x (L + \mathbf{J}_\mu \cdot \mathbf{A}_\mu) \right] \tag{2.226}$$

Here we set the time component of the external source to zero $J_0 = 0$, by definition.

Recall that

$$M_c^{ab}(x, y) = \nabla^2 \left[ \delta^{ab} \delta^3(x - y) + g\epsilon^{abc} G_0(x, y) A_i^c(y) \nabla_i \right] \delta(x_0 - y_0) \tag{2.227}$$

so that

$$\det M_c = \det \nabla^2 \det(1 + L) \tag{2.228}$$

$$1^{ab} = \delta^{ab} \delta^4(x - y) \tag{2.229}$$

$$L^{ab} = g\epsilon^{abc} G_0(x, y) A_i^c(y) \nabla_i \delta(x_0 - y_0) \tag{2.230}$$

We may then evaluate

$$\det(1 + L) = \exp\left[ \mathrm{Tr} \log(1 + L) \right] \tag{2.231}$$

$$= \exp \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}$$

$$\cdot \int d^4x_1 \cdots d^4x_n \, \mathrm{Tr}\big( L(x_1, x_2) L(x_2, x_3) \cdots L(x_n, x_1) \big) \tag{2.232}$$

This term modifies the Feynman rules at each order and has appearance of a closed-loop expression. We shall see below, in a more general treatment, how this modification is incorporated correctly.

At this point it is interesting to note that the correct Feynman rules can be extracted from Eqs. (2.226), (2.228), and (2.232) and it is a little surprising that no one did such an analysis between the invention of Yang–Mills theory in 1954 and the works of Faddeev and Popov [18–20] in 1967 and, specific to the Coulomb gauge, the work of Khriplovich [15] in 1970.

Before proceeding to the general treatment (for any gauge) of Faddeev and Popov, there is one subtlety to tidy up in the analysis above, as noted following Eq. (2.197). The point is that the operator

$$O^{ab}\big( \mathbf{A}_i(x) \big) = D_k^{ab} \nabla_k \tag{2.233}$$

is singular (i.e., has zero determinant), and hence no Green's function satisfying Eq. (2.197) exists [16, 17].

The way to proceed in this situation is to enumerate the nonzero eigenvalues and the zero modes according to ($M$ and $N$ may be infinite)

$$O^{ab}(x)\psi^{b(n)}(x) = \lambda^{(n)}\psi^{a(n)}(x) \quad n = 1, 2, \ldots, N, \quad \lambda^{(n)} \neq 0 \tag{2.234}$$

$$O^{ab}(x)h^{b(m)}(x) = 0 \qquad m = 1, 2, \ldots, M \tag{2.235}$$

The completeness relation now reads

$$\sum_{n=1}^{N} \psi_a^{(n)}(x)\psi_b^{(n)*}(y) \sum_{m=1}^{M} h_a^{(m)}(x)h_b^{(m)*}(y) = \delta_{ab}\delta(x-y) \tag{2.236}$$

We may conveniently divide the vector space $F$ spanned by the eigenvectors of $O^{ab}(x)$ into two orthogonal subspaces: the spurious subspace $S$ spanned by the zero modes $h_a^{(m)}(x)$ and the physical subspace $P$ spanned by the $\psi_a^{(n)}(x)$. Projecting on the subspace $P$ then allows us to define a unique inverse $\mathcal{G}$ acting on $P$ of $O^{ab}(x)$ satisfying

$$O^{ab}\big(\mathbf{A}_i(x)\big)\mathcal{G}^{bc}(x, y, A) = \delta^{ac}\delta(x-y) - \sum_{m=1}^{M} h_a^{(m)}(x)h_c^{(m)*}(y) \tag{2.237}$$

Let us define

$$\bar{O}^{ab}(x) = -g\epsilon^{abc}A_i^c(x)\nabla_x^i \tag{2.238}$$

so that

$$O^{ab}(x) = \delta^{ab}\nabla_x^2 + \bar{O}^{ab}(x) \tag{2.239}$$

Equation (2.237) may now be recast in the integral form

$$\begin{aligned}\mathcal{G}_{ab}(x, y, A) \;=\; \delta_{ab}G_0(x, y) &- \int d^3z G_0(x, z) \sum_m h_a^{(m)}(z)h_b^{(m)*}(y) \\ &- \int d^3z G_0(x, z)\bar{O}^{ac}(z)\mathcal{G}_{cb}(z, y, A)\end{aligned} \tag{2.240}$$

where $G_0(x, y)$ is the Abelian Coulomb Green's function, which satisfies Eq. (2.219).

To confirm Eq. (2.240) is immediate. Suppressing all internal symmetry labels, and integrations, one has

$$\mathcal{G} = G_0(\nabla^2 + \bar{O})\mathcal{G} - G_0\overline{O}\mathcal{G} \tag{2.241}$$

$$= G_0 O G - G_0\overline{O}\mathcal{G} \tag{2.242}$$

$$= G_0\left(1 - \sum hh^*\right) - G_0\overline{O}\mathscr{G} \tag{2.243}$$

$$= G_0 - G_0\sum hh^* - G\overline{O}\mathscr{G} \tag{2.244}$$

as required.

Now, the Green's function (2.240) differs from that of the less careful earlier analysis only by the presence of the extra term containing zero modes. But the density $\nabla^2\mathbf{f}$ of Eq. (2.197) lies[3)] in the physical subspace $P$ and hence is orthogonal to all vectors in $S$, that is,

$$\left(\nabla^2\mathbf{f}\cdot\mathbf{h}^{(m)}\right) = 0 \quad \text{for all } m \tag{2.245}$$

and hence

$$\mathscr{G}\nabla^2\mathbf{f} = G\nabla^2\mathbf{f} \tag{2.246}$$

This has the consequence that when we use the correct Green's function, Eq. (2.240), in perturbation theory, then since $\mathscr{G}$ acts always on vectors in the physical subspace $P$, the zero-mode term gives no new contribution and the results are unchanged from those derived earlier.

It should be noted, however, that although the zero modes do not contribute terms analytic at $g = 0$, they may provide important nonperturbative effects of order $1/g$ or $\exp(a/g)$. These may be dominant in, for example, quantum chromodynamics [21–23], where the expected spectrum of bound-state hadrons is not visible at any order of the perturbation expansion in $g$.

The canonical quantization above has been restricted to the Coulomb gauge, and we have arrived at the generating functional,

$$\begin{aligned} W_c[J] = &\int D\mathbf{A}_\mu\, D\mathbf{F}_{\mu\nu} \det M_c \prod_x \delta(\nabla_i A_i) \\ &\cdot \exp\left[i\int d^4x(L + \mathbf{J}_\mu\cdot\mathbf{A}_\mu)\right] \end{aligned} \tag{2.247}$$

We cannot do the $\mathbf{F}_{\mu\nu}$ path integral to arrive at the second-order form

$$\begin{aligned} &W_C[J] \\ &= \int D\mathbf{A}_\mu\left[\det M_c \prod_x \delta(\nabla_i\mathbf{A}_i)\right] \\ &\quad\cdot \exp\left\{i\int d^4x\left[-\frac{1}{4}(\partial_\mu\mathbf{A}_\nu - \partial_\nu\mathbf{A}_\mu + g\mathbf{A}_\mu{}^\wedge\mathbf{A}_\nu)^2 + \mathbf{J}_\mu\cdot\mathbf{A}_\mu\right]\right\} \end{aligned} \tag{2.248}$$

3) This statement [and Eq. (2.245)] is really a physical assumption and we do not claim mathematical rigor here.

Except for the term in the first set of bracket, on the right side, this is the standard form in Section 2.2. So why is the standard form not applicable when there is local gauge invariance? Following Faddeev and Popov [18–20] (see also Fradkin and Tyutin [24, 25]), we may express the reason in at least two equivalent forms:

1. The propagator is the inverse of the matrix in the quadratic piece of the Lagrangian. This is unique for a nongauge theory, but in a gauge theory, the matrix is singular. From Eq. (2.248) the quadratic term is

$$\frac{1}{2}A_\mu(x)K^{\mu,\nu}(x,y)A_\nu(y) \tag{2.249}$$

   with

$$K^{\mu\nu}(x,y) = -\left(\nabla^2 g^{\mu\nu} - \partial_\mu\partial_\nu\right)\delta^4(x-y) \tag{2.250}$$

   This has zero eigenvalues ($K \cdot A^L = 0$), and hence the propagator is not defined. Because of local gauge invariance, the longitudinal part $A_i^L$ in the path integral is completely undamped.
2. An even more useful viewpoint is to observe that within the manifold of all fields $\mathbf{A}_\mu(x)$ encountered in the domain of the path integral, the action $S$ is constant on orbits of the gauge group because of local gauge invariance. What is required is evaluation of the functional only for *distinct* orbits, and it seems reasonable [18] to divide out the infinite orbit volume.

In this spirit we define a hypersurface within the manifold of all fields by the gauge-fixing condition

$$F_a(\mathbf{A}_\mu) = 0 \quad a = 1, 2, \ldots, N \tag{2.251}$$

where $N$ is the number of generators of the gauge group. Let us denote a gauge function by $g$; then (with $A_\mu = A_\mu^a T^a$)

$$A_\mu^g = U(g)A_\mu U^{-1}(g) - \frac{i}{g}\left(\partial_\mu U(g)\right)U^{-1}(g) \tag{2.252}$$

The equation

$$F_a\left(\mathbf{A}_\mu^g\right) = 0 \tag{2.253}$$

is assumed to have one unique solution $g$ so that any orbit intersects this hypersurface precisely once. This is always true locally within the group space, although gauge-fixing degeneracy [16] leads globally to multiple intersections [26, 27], which are handled later.

Let the unitary gauge transformation operator be $U(g)$ as in Eq. (2.252); that is,

$$U(g) = \exp\left[i\mathbf{T}\cdot\mathbf{g}(x)\right] \tag{2.254}$$

Then

$$U(g)U(g') = U(gg') \tag{2.255}$$

and

$$d(g') = d(gg') \tag{2.256}$$

follow from the group properties.

Writing

$$U(g) \simeq 1 + i\mathbf{g}\cdot\mathbf{T} + O\big(g^2\big) \tag{2.257}$$

we may express

$$dg \equiv \prod_a dg_a(x) \tag{2.258}$$

for infinitesimal transformations $g \simeq 1$.

The argument of Faddeev and Popov [18] proceeds as follows. We begin with the naive expression for the vacuum-to-vacuum amplitude

$$W[0] = \int D\mathbf{A}_\mu \exp\left(i\int d^4x L\right) \tag{2.259}$$

Now define $\Delta_F[\mathbf{A}_\mu]$ such that

$$\Delta_F[\mathbf{A}_\mu]\int \prod_x dg(x) \prod_{x,a} \delta\big(F^a\big(\mathbf{A}_\mu^g(x)\big)\big) = 1 \tag{2.260}$$

The quantity $\Delta_F[\mathbf{A}_\mu]$ is easily shown to be gauge invariant using

$$\Delta_F^{-1}[\mathbf{A}_\mu] = \int \prod_x dg'(x) \prod_{x,a} \delta\big(F^a\big(A_\mu^{g'g}(x)\big)\big) \tag{2.261}$$

$$= \int \prod_x d\big(g(x)g'(x)\big) \prod_{x,a} \delta\big(F^a\big(A_\mu^{g'g}(x)\big)\big) \tag{2.262}$$

$$= \int \prod_x dg''(x) \prod_{x,a} \delta\big(F^a\big(A_\mu^{g''}(x)\big)\big) \tag{2.263}$$

$$= \Delta_F^{-1}[\mathbf{A}_\mu] \tag{2.264}$$

We may insert the resolution of the identity (2.260) into the naive expression (2.259) to arrive at

$$W[0] = \int \prod_x dg(x) \int DA_\mu \Delta_F[\mathbf{A}_\mu] \prod_{x,a} \delta\big(F^a(\mathbf{A}_\mu^g)\big) \cdot \exp\left[i \int d^4x L(x)\right] \tag{2.265}$$

Using the gauge invariance of $\Delta_F$ and $L$, the functional integral may be considered as an integral over $A_\mu^g$, and we may drop the infinite factor

$$\int \prod_x dg(x) \tag{2.266}$$

and define

$$W[\mathbf{J}] = \int D\mathbf{A}_\mu \Delta_F[\mathbf{A}_\mu] \prod_{x,a} \delta\big(F^a(\mathbf{A}_\mu(x))\big) \cdot \exp\left\{i \int d^4x \big[L(x) + \mathbf{A}_\mu J_\mu\big]\right\} \tag{2.267}$$

Our next task is the computation of $\Delta_F$. Let us define the matrix $M_F$ by

$$F^a\big(\mathbf{A}_\mu^g(x)\big) = F^a\big(\mathbf{A}_\mu(x)\big) + \frac{1}{g} \int d^4y \sum_b M_F(x,y)_{ab} g_b(y) + O\big(g^2\big) \tag{2.268}$$

Then the Faddeev–Popov determinant is

$$\Delta_F^{-1}[A_\mu] = \int \prod_{x,a} dg_a(x) \delta\big(F^a(\mathbf{A}_\mu^g(x))\big) \tag{2.269}$$

$$= \int \prod_{x,a} \left[dg_a(x) \delta\left(\frac{1}{g} M_F g\right)\right] \tag{2.270}$$

so that, within the constant of proportionality $(1/g)$,

$$\Delta_F[\mathbf{A}_\mu] = \det(M_F) \tag{2.271}$$

$$= \exp(\mathrm{Tr} \ln M_F) \tag{2.272}$$

Let us compute this in certain examples:

**Example 1.** Coulomb gauge

$$F^a(\mathbf{A}_\mu) = \nabla_k A_k^a \tag{2.273}$$

Note that

$$U(g) = 1 + i\mathbf{g}\cdot\mathbf{T} \tag{2.274}$$

$$\mathbf{A}_\mu^g \cdot \mathbf{T} = U(g)\left[\mathbf{A}_\mu \cdot \mathbf{T} + \frac{1}{ig}U^{-1}(g)\partial_\mu U(g)\right]U^{-1}(g) \quad \left[\text{cf. Eq. (1.71)}\right] \tag{2.275}$$

$$= (1 + ig_a T_a)\left[A_\mu^b T^b + \frac{1}{ig}\left(1 - ig^c T^c\right)iT^d\partial_\mu g^d\right]\left(1 - ig^e T^e\right) \tag{2.276}$$

$$= A_\mu^a T_a + ig_a\left[T^a, T^b\right]A_\mu^b + \frac{1}{ig}\left(iT^d\partial_\mu g^d\right) \tag{2.277}$$

$$= A_\mu^a T^a + \frac{1}{g}T^a\partial_\mu g^a - \epsilon^{abc}A_\mu^b g_b T^a + O\left(g^2\right) \tag{2.278}$$

This is for any gauge; specializing to the Coulomb gauge yields

$$\nabla_i \mathbf{A}_i^g = \nabla_i \mathbf{A}_i + \frac{1}{g}\left(\nabla^2\delta^{ab} - g\epsilon^{abc}A_i^c\nabla_i\right)g_b \tag{2.279}$$

and hence the required determinant is of the matrix

$$M_C^{ab}(x, y) = \left(\nabla^2\delta^{ab} - g\epsilon^{abc}A_i^c\nabla_i\right)\delta^4(x - y) \tag{2.280}$$

This agrees precisely with the result of canonical quantization given earlier, in Eq. (2.227).

**Example 2.** Lorenz gauge (also called the Landau gauge)

$$F^a(\mathbf{A}_\mu) = \partial_\mu A_\mu^a \tag{2.281}$$

From Eq. (2.278) we find that

$$\partial_\mu\left(A_\mu^g\right)^a = \partial_\mu A_\mu^a + \frac{1}{g}\left(\partial^2\delta^{ab} - g\epsilon^{abc}A_\mu^c\partial_\mu\right)g_b + O\left(g^2\right) \tag{2.282}$$

and consequently,

$$M_L^{ab}(x, y) = \left(\partial^2\delta^{ab} - g\epsilon^{abc}A_\mu^c\partial_\mu\right)\delta^4(x - y) \tag{2.283}$$

for Landau gauge.

**Example 3.** Axial gauge

$$F^a(A_\mu) = n_\mu A_\mu^a \tag{2.284}$$

$$n_\mu\left(A_\mu^g\right)^a = n_\mu \mathbf{A}_\mu^a + \frac{1}{g}\left(n_\mu\partial_\mu\delta^{ab} - g\epsilon^{abc}n_\mu A_\mu^c\right) + O\left(g^2\right) \tag{2.285}$$

Restricting $\mathbf{A}_\mu$ to those respecting $n_\mu \mathbf{A}_\mu = 0$, we see that for the axial gauge the Faddeev–Popov determinant is independent of the gauge field $\mathbf{A}_\mu$ and hence can be omitted as an irrelevant overall factor. This means that for the axial gauge $n_\mu \mathbf{A}_\mu = 0$, no ghost loops (see below) occur. This remark includes the special cases of the lightlike gauge ($n^2 = n_\mu n_\mu = 0$) and the Arnowitt–Fickler [28] gauge ($n_\mu = \delta_{\mu 3}$). The awkwardness of axial gauge is that the Feynman rules involve $n_\mu$ explicitly and is thus not explicitly Lorentz invariant.

## 2.4
## Feynman Rules

The Feynman rules for non-Abelian theories are gauge dependent. We shall focus on the covariant gauge $\partial_\mu \mathbf{A}_\mu = 0$, since in a general gauge, manifest Lorentz invariance is lost. For the Lorentz gauge we have arrived at

$$W_L[\mathbf{J}] = \int D\mathbf{A}_\mu \prod_x \delta\big(\partial_\mu \mathbf{A}_\mu(x)\big) \det(M_L) \exp\left[i \int d^4x (L + \mathbf{J}_\mu \cdot \mathbf{A}_\mu)\right] \tag{2.286}$$

To find the perturbation theory rules, we need to promote the delta function and determinant factors into the exponential to arrive at

$$W_L[0] \simeq \int DA_\mu \exp\big[i(S + S_{\mathrm{GF}} + S_{\mathrm{FPG}})\big] \tag{2.287}$$

where the subscripts refer to gauge-fixing and Faddeev–Popov ghost terms.

For the delta function we use the formula

$$\prod_x \delta\big(\partial_\mu \mathbf{A}_\mu(x)\big) = \lim_{\alpha \to 0} \exp\left[-\frac{1}{2\alpha} \int d^4x (\partial_\mu A_\mu)^2\right] \tag{2.288}$$

The determinant factor gives rise to ghost loops and is handled as follows. We have

$$\det(M_L) = \exp(\mathrm{Tr} \ln M_L) \tag{2.289}$$

$$M_L = \big(\partial^2 + g\partial^\mu \mathbf{A}_\mu{}^\wedge\big) \tag{2.290}$$

Now introduce

$$S_{\mathrm{FPG}} = \int d^4x \big(\partial_\mu \mathbf{c}^+ \cdot \partial_\mu \mathbf{c} + g\partial_\mu \mathbf{c}^+ \cdot \mathbf{A}_\mu{}^\wedge \mathbf{c}\big) \tag{2.291}$$

$$= \int d^4x\, d^4y \sum_{a,b} c_a^+(x) M_L^{ab}(x, y) c_b(y) \tag{2.292}$$

where $\mathbf{c}$ is a complex adjoint representation of scalar fields.

With normal commutation relations for **c**, one would have

$$\int D\mathbf{c}\, D\mathbf{c}^{+} e^{iS_{\text{FPG}}} \sim \left[\det(M_L)\right]^{-1} \tag{2.293}$$

To obtain the determinant in the numerator as required, rather than the denominator, it is necessary that the **c** fields satisfy anticommutation relations, because then

$$\int Dc\, Dc^{+} e^{iS_{\text{FPG}}} \sim \det(M_L) \tag{2.294}$$

The reason for this can be understood simply if we write [cf. Eq. (2.283)]

$$M_L = \nabla^2(1 + L) \tag{2.295}$$

whereupon (for commutation relations)

$$\left[\det(M_L)\right]^{-1} \sim \exp\left[-\operatorname{Tr}\ln(1 + L)\right] \tag{2.296}$$

$$= \exp\left[\sum_n \frac{(-1)^n}{n} \operatorname{Tr} L^n\right] \tag{2.297}$$

For anticommutation relations, each trace switches sign, as in a closed loop, and one has instead

$$\exp\left[\sum_n \frac{(-1)^{n+1}}{n} \operatorname{Tr} L^n\right] = \det(M_L) \tag{2.298}$$

To summarize, we now have

$$W_L[\mathbf{J}] = \int D\mathbf{A}_\mu D\mathbf{c}\, D\mathbf{c}^{+} \exp\left[i(S + S_{\text{GF}} + S_{\text{FPG}}) + i \int d^4x \mathbf{J}_\mu \cdot \mathbf{A}_\mu\right] \tag{2.299}$$

where

$$S = \int d^4x L(x) \tag{2.300}$$

$$S_{\text{GF}} = -\frac{1}{2\alpha} \int d^4x (\partial_\mu A_\mu)^2 \tag{2.301}$$

$$S_{\text{FPG}} = +\int d^4x \left(\partial_\mu c_a^{+} \partial_\mu c_a + g\epsilon_{abc}\partial_\mu c_a^{+} A_\mu^b c_c\right) \tag{2.302}$$

so the effective Lagrangian is

$$L_{\text{eff}} = -\frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a - \frac{1}{2\alpha}(\partial_\mu A_\mu)^2 + \left(\partial_\mu c_a^{+} \partial_\mu c_a + g\epsilon_{abc}\partial_\mu c_a^{+} A_\mu^b c_c\right) \tag{2.303}$$

with the limit $\alpha \to 0$ understood.

To find the propagator we set $g = 0$, whereupon

$$W_L^0[J]$$

$$= \lim_{\alpha\to 0} \int D\mathbf{A}_\mu \exp\left\{ i \left[ \int d^4x A_\mu^a(x) \right.\right.$$

$$\left.\left. \cdot \left[ -\partial^2 g_{\mu\nu} + \partial_\mu \partial_\nu \left( 1 - \frac{1}{\alpha} \right) \right] A_\nu^a(x) + \int d^4x J_\mu^a(x) A_\mu^a(x) \right] \right\} \tag{2.304}$$

$$= \lim_{\alpha\to 0} \exp\left[ -\frac{1}{2} \int d^4x\, d^4y J_\mu^a(x) D_{\mu\nu}^{ab}(x-y) J_\nu^b \right] \tag{2.305}$$

with

$$D_{\mu\nu}^{ab}(x-y) = \delta^{ab} \int \frac{d^4p}{(2\pi)^4} \frac{e^{-ip(x-y)}}{p^2 + i\epsilon} \left[ -g_{\mu\nu} + (1-\alpha) \frac{p_\mu p_\nu}{p^2} \right] \tag{2.306}$$

So far, we have proved that this propagator is valid only for $\alpha = 0$ (the Landau gauge). It is, however, straightforward to prove that it is true for any value of $\alpha$, as follows.

We adopt the slightly more general gauge

$$\partial_\mu \mathbf{A}_\mu(x) = \mathbf{f}(x) \tag{2.307}$$

where $\mathbf{f}(x)$ is an arbitrary function. Then since the path integral is independent of $\mathbf{f}(x)$, we may include an arbitrary functional, $G[F]$, to rewrite Eq. (2.286) as

$$W[\mathbf{J}] = \int D\mathbf{f}\, D\mathbf{A}_\mu \prod_x \delta\big(\partial_\mu \mathbf{A}_\mu(x) - \mathbf{f}(x)\big) \det(M_L)$$

$$\cdot \exp\left[ i \int d^4x (L + \mathbf{J}_\mu \cdot \mathbf{A}_\mu) \right] G(\mathbf{f}) \tag{2.308}$$

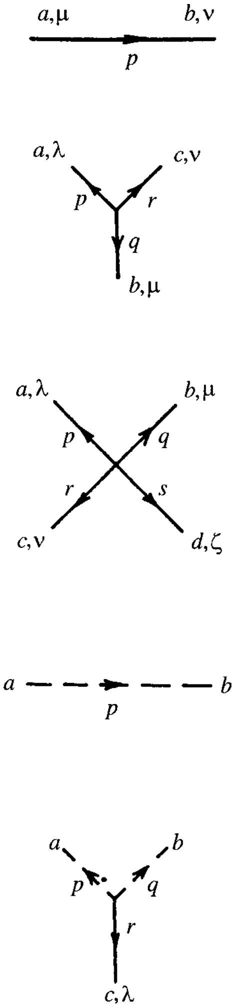
By making the choice

$$G(\mathbf{f}) = \exp\left( -\frac{i}{2\alpha} \int d^4x \mathbf{f}^2 \right) \tag{2.309}$$

we find that

$$W[J] = \int D\mathbf{A}_\mu \det(M_L) \exp\left\{ i \int d^4x \left[ L - \frac{1}{2\alpha} (\partial_\mu \mathbf{A}_\mu)^2 + \mathbf{J}_\mu \mathbf{A}_\mu \right] \right\} \tag{2.310}$$

and this leads to the propagator, Eq. (2.306), for general $\alpha$. With $\alpha = 1$ it is the Feynman gauge that is valuable for diagrams with internal vector propagators since the number of terms is reduced.




$$-i\delta^{ab}[g_{\mu\nu} - (1-\alpha)p_\mu p_\nu/p^2]/p^2$$

$$-\epsilon^{abc}[(p-q)_\nu g_{\lambda\mu} + (q-r)_\lambda g_{\mu\nu} + (r-p)_\mu g_{\lambda\nu}]$$

$$\begin{aligned}&- i\epsilon^{abf}\epsilon^{cdf}(g_{\lambda\nu}g_{\mu\zeta} - g_{\lambda\zeta}g_{\mu\nu})\\ &- i\epsilon^{acf}\epsilon^{bdf}(g_{\lambda\mu}g_{\nu\zeta} - g_{\lambda\zeta}g_{\mu\nu})\\ &- i\epsilon^{adf}\epsilon^{cbf}(g_{\lambda\nu}g_{\mu\zeta} - g_{\lambda\mu}g_{\zeta\nu})\end{aligned}$$

$$i\delta^{ab}/p^2$$

$$\epsilon^{abc}p_\lambda$$

**Figure 2.6** Feynman rules (all momenta outgoing) for pure Yang–Mills theory in Landau gauge ($\alpha = 0$) and Feynman gauge ($\alpha = 1$). As shown in the text, these covariant rules are valid for any value of the parameter $\alpha$.

The ghost propagator is similarly evaluated, and the rules for vertices follow by turning on the coupling strength $g \neq 0$. The final results for the SU(2) Feynman rules in the covariant gauges are collected together in Fig. 2.6.

The scalar particles represented by the fields $\mathbf{c}$ and $\mathbf{c}^+$ are purely fictitious and do not occur as asymptotic states (they would violate the necessary spin-statistics theorem) nor as poles of the physical $S$-matrix. Similar closed-loop ghosts can be shown to be necessary in general relativity; one difference is that there the ghosts carry spin 1.

Concerning the rules tabulated in Fig. 2.6, note that the ghost-vector vertex is dashed on the ghost line, which is differentiated; a consistent rule is then that a given ghost propagator must never be dashed at both ends. Also note that because of its fermionic nature, a closed scalar loop gives a sign change $(-1)$.

The Feynman rules (Fig. 2.6) will be used in our later discussions of perturbative renormalizability and of asymptotic freedom. Some preliminary remarks on renormalizability are in order here. Because the vertices are all of dimension 4, and more especially because the vector propagator, for any value of $\alpha$, has high-energy behavior no worse than a scalar propagator, the symmetric Yang–Mills theory is renormalizable by the usual power-counting arguments [29–31]. (Reference [29] is reprinted in Ref. [4].) For suppose that a diagram contains

| | |
|---|---|
| $n_3$ | Three-vector vertices |
| $n_4$ | Four-vector vertices |
| $n_g$ | Ghost-vector vertices |
| $E$ | External vectors |
| $I_v$ | Internal vector propagators |
| $I_g$ | Internal ghost propagators |

then the superficial degree of divergence is

$$D = n_3 + n_g + 2I_v + 2I_g - 4(n_3 + n_4 + n_g) + 4 \tag{2.311}$$

Now use the identities

$$2I_g = 2n_g \tag{2.312}$$

$$2I_v + E = 3n_3 + 4n_4 + ng \tag{2.313}$$

to find

$$D = 4 - E \tag{2.314}$$

so that the number of renormalizable parts is finite, being limited to $E \leqslant 4$. Of course, this does not prove renormalizability, which requires a careful regularization procedure that respects gauge invariance and the generalized Ward identities; this was first carried through by 't Hooft [32].

This is for the symmetric case. When there is spontaneous breaking with a Higgs mechanism, the vector propagator would normally acquire a $p_\mu p_\nu/M^2$ term, which would lead to the replacement $2I_v$ by $4I_v$ in Eq. (2.311) and hence failure of the power-counting argument. It was shown by 't Hooft [33] that there exists a judicious choice of gauge where the mischievous term $(p_\mu p_\nu/M^2)$ is absent in the Feynman rules; 't Hooft [33] was then also able to demonstrate renormalizability for the spontaneously broken case. More details of these statements about perturbative renormalizability are provided in Chapter 3.

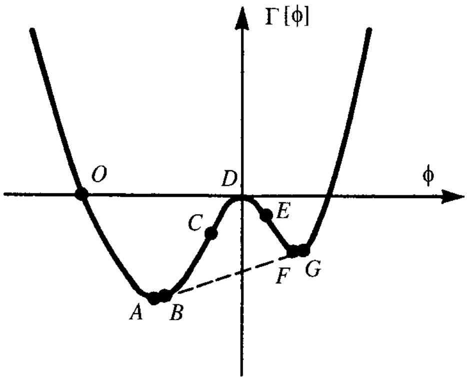



**Figure 2.7** $\Gamma - \phi$ plot.

Note that these remarks apply to renormalization of the ultraviolet infinities. Handling of the infrared divergences, arising from the presence of massless particles, is a completely separate question.

## 2.5 Effects of Loop Corrections

So far we have considered the potential function $V(\phi)$ only in the tree approximation. It turns out that in several important cases the one-loop contribution[4] to the potential can be evaluated and can significantly alter the conclusions concerning the presence or absence of spontaneous symmetry breaking.

It will be important in the specific examples to be discussed—scalar electrodynamics, an $O(N)$ model, and the Glashow–Salam–Weinberg model—that the true ground state is a global minimum of the effective potential; a local nonglobal minimum is a candidate only for a metastable ground state. Unstable vacua are themselves interesting and important and are discussed at the end of this section.

Suppose that the functional $\Gamma[\phi]$ has two unequal minima, as indicated in Fig. 2.7. Since

$$\frac{\partial \Gamma}{\partial \phi} = -J \tag{2.315}$$

the $J - \phi$ plot appears as indicated in Fig. 2.8. Assume that this $\Gamma[\phi]$ is the result of evaluating the effective potential (at zero external momenta) at the tree approximation, or possibly including the contributions from loop diagrams up to some finite number of loops.

4) It is interesting to bear in mind here that the loop expansion corresponds to the semiclassical expansion in $h$ (see Refs. [34] and [35]).

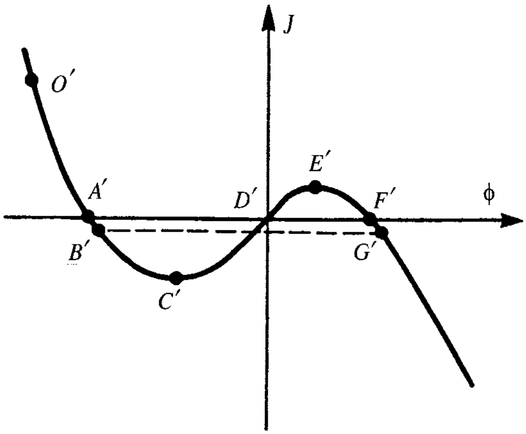



**Figure 2.8** $J-\phi$ plot.

The exact $\Gamma[\phi]$, computed from the full theory, must satisfy a convexity property as follows. The inverse [one-particle-irreducible (1PI) propagator is given by

$$\frac{\delta^2\Gamma}{\delta\phi^2} = \Delta^{-1} = m^2 \geqslant 0 \tag{2.316}$$

and must be positive definite if there are no tachyons. Thus, the exact $\Gamma[\phi]$ must be convex downward and cannot have the form indicated in Fig. 2.7. The segment in the vicinity of the unstable local maximum (*CDE* in Fig. 2.7 violates this condition and thus must deviate from the exact $\Gamma[\phi]$. The corresponding segment in a $J-\phi$ plot, of course, violates the corresponding condition,

$$\frac{\delta J}{\delta\phi} \geqslant 0 \tag{2.317}$$

There are segments between the minima (*AC*, *EF*) which respect the convexity property but which do not always correspond to accessible, completely stable situations.

It is fruitful to regard the two minima $A$ and $F$ as corresponding to two possible phases of the theory. Although, in the approximation to $\Gamma[\phi]$ considered, $J$ vanishes at both phases, only one of them—the global minimum—is accessible when we take the limit $J \to 0$ appropriately, either from positive or negative values of $J$.

While varying the external source J, we require that the energy density always be minimal. Taking as a reference $\Gamma[\phi_0] = 0$ as indicated in Fig. 2.7 the free energy is given by

$$-\int_{\phi_0}^{\phi} J\, d\phi = -\int_{\phi_0}^{\phi} \frac{\delta\Gamma}{\delta\phi}\delta\phi \tag{2.318}$$

$$= \Gamma[\phi] \tag{2.319}$$

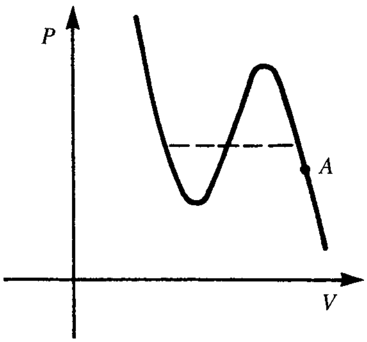



**Figure 2.9** $P-V$ isotherm.

We are considering continuous variations of the external source $J$, and therefore $\delta\Gamma/\delta\phi$ must be continuous in $\phi$. For some value of $J$, however, the system will make (at constant $J$) a first-order transition between the two phases.

To represent this situation in terms of the calculated $\Gamma[\phi]$, the only possibility is to draw the tangent line $BG$ indicated in Fig. 2.7. For this line

$$\Gamma[\phi_G] - \Gamma[\phi_B] = -\int_{\phi_B}^{\phi_G} J\, d\phi \tag{2.320}$$

$$= J_B(\phi_B - \phi_G) \tag{2.321}$$

and hence corresponds to a rule of equal areas on the $J-\phi$ plot—the areas above and below the line are equal.

As $J$ is varied continuously, therefore, one reaches the point $\phi_G$ on the original curve; then there is a transition at constant $J$ to the point $\phi_B$. Further increasing $J$, one reaches $\phi_A$ for vanishing $J$. Thus only the global minimum $A$ corresponds to a stable ground state of the theory; the nonglobal minimum $F$ is never reached, since it does not correspond to lowest energy.

For more than two minima, it is straightforward to extend these arguments, including the equal-area rule; again, only a global minimum corresponds to a completely stable ground state.

The situation in quantum field theory has an analogy in statistical mechanics if we identify $J$ and $\phi$ with appropriate thermodynamic coordinates. In a liquid–gas transition, the identification would be with pressure and volume, respectively, and the approximate form of $\Gamma[\phi]$ then corresponds to an approximation to the Helmholtz free energy $H$. For example, the van der Waals equation of state gives rise to a typical isotherm as indicated in Fig. 2.9; this violates the convexity property

$$\frac{\partial P}{\partial V} \leqslant 0 \tag{2.322}$$

which must hold for the exact solution. By the Maxwell construction, which minimizes $H$ while $P$ is varied, one obtains a solution that respects the required convexity. Note that the global minimum in the earlier discussion would correspond here to a point $A$ on the isotherm (Fig. 2.9).

In the examples we are about to consider, the fundamental assumption is that provided that the coupling constants are sufficiently small, the calculation of $\Gamma[\phi]$ from tree diagrams and one-loop corrections gives an accurate estimate except for the region of the phase transition. In particular, the assertion is that the location of the global minimum ($A$) is accurately given; it should be emphasized, however, that this assertion, although plausible, has not been vigorously justified. For the transition region, of course, the approximation cannot reproduce the exact nonanalytic behavior and gives instead some analytic continuation of the exact $\Gamma[\phi]$.

**Example 1.** In this example of one-loop corrections, we consider the Lagrangian density $L$ given by

$$L = -\frac{1}{4}F_{\mu\nu}F_{\mu\nu} + D_\mu\phi^* D_\mu\phi - m_0^2\phi^*\phi - \frac{\lambda}{6}(\phi^*\phi)^2 \tag{2.323}$$

where

$$D_\mu\phi = (\partial_\mu + ieA_\mu)\phi \tag{2.324}$$

Putting $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$, this becomes

$$\begin{aligned} L = {} & \frac{1}{2}(\partial_\mu\phi_1 - eA_\mu\phi_2)^2 + \frac{1}{2}(\partial_\mu\phi_2 + eA_\mu\phi_1)^2 \\ & - \frac{1}{2}m_0^2(\phi_1^2 + \phi_2^2) - \frac{\lambda}{4!}(\phi_1^2 + \phi_2^2)^2 - \frac{1}{4}F_{\mu\nu}F_{\mu\nu} \end{aligned} \tag{2.325}$$

When $m_0^2 > 0$, this is scalar electrodynamics and the normal perturbative solution holds for small couplings. When $m_0^2 < 0$, there is spontaneous breakdown and the Higgs mechanism occurs. When $m_0^2 = 0$, which vacuum does the theory pick?

To answer this question, we compute the one-loop contribution to the effective potential [36]. Since $V$ is a function only of $(\phi_1^2 + \phi_2^2)$, we need consider only $\phi_1$ as external legs. The relevant graphs are indicated in Fig. 2.10. First consider Fig. 2.10$a$, where $\phi_1$ is circulating in the loop and the relevant coupling is $-\lambda\phi_1^4/4!$. These contributions to $V$ are given by the sum (recall that external momenta are set equal to zero in evaluating $V$):

$$V = i\int \frac{d^4k}{(2\pi)^4}\sum_{n=1}^{\infty}\frac{1}{2n}\frac{(1/2\lambda\phi_1^2)^n}{(k^2+i\epsilon)^n} \tag{2.326}$$

Some explanation of this will be helpful:

1. The Feynman rules are $-i\lambda$ at each vertex and $i/(k^2 + i\epsilon)$ for each propagator, so the $i$ factors combine to give $(-i^2)^n = 1$.
2. The overall $i$ is from the definition $V \sim iS$, where $S$ is the $S$-matrix.
3. The $1/2^n$ is from Bose symmetrization of the two external $\phi_1$ legs at each vertex.
4. The $1/2n$ allows for the fact that cyclic and anticyclic permutations of the vertices do not alter the diagram.

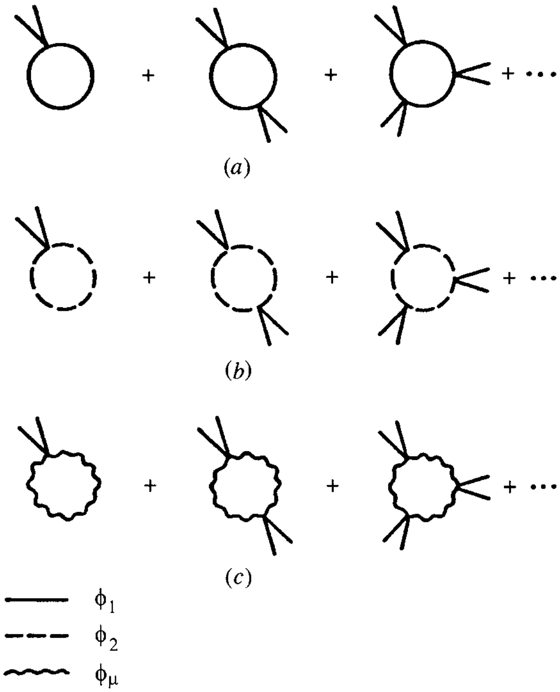



**Figure 2.10** One loop graphs for scalar electrodynamics: (*a*) $\phi_1$ loop; (*b*) $\phi_2$ loop; (*c*) $A_\mu$ loop.

Rotating to Euclidean space, using $i\,dk_0 \to -dk_4$ and $k^2 \to -k^2$, gives

$$V = \frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\ln\left(1 + \frac{\lambda\phi_1^2}{2k^2}\right) \tag{2.327}$$

Cutting off the (divergent) integral at $k^2 = \Lambda^2$ and changing the variable to $y = \lambda\phi_1^2/2k^2$, one obtains

$$V = \frac{\lambda^2\phi_1^4}{128\pi^2}\int_{\lambda\phi_1^2/2\Lambda^2}^{\infty}\frac{dy}{y^3}\ln(1+y) \tag{2.328}$$

where we used

$$\int d^4k = 2\pi^2\int k^3\,dk = \pi^2\int k^2\,d(k)^2 \tag{2.329}$$

Now use the definite integral

$$\int_{y_0}^{\infty}\frac{dy}{y^3}\ln(1+y) = \frac{\ln(1+y_0)}{2y_0^2} + \frac{1}{2y_0} - \frac{1}{2}\ln\left(1+\frac{1}{y_0}\right) \tag{2.330}$$

to find

$$V = \frac{\lambda^2\phi_1^4}{128\pi^2}\left[\frac{2\Lambda^4}{\lambda^2\phi_1^4}\ln\left(1+\frac{\lambda\phi_1^2}{2\Lambda^2}\right)+\frac{\Lambda^2}{\lambda\phi_1^2}-\frac{1}{2}\ln\left(1+\frac{2\Lambda^2}{\lambda\phi_1^2}\right)\right] \tag{2.331}$$

Expanding the logarithms and dropping terms that vanish for large cutoff gives

$$V = \frac{\lambda\Lambda^2\phi_1^2}{64\pi^2}+\frac{\lambda^2\phi_1^2}{256\phi^2}\left(\ln\frac{\lambda\phi^2}{2\Lambda^2}-\frac{1}{2}\right) \tag{2.332}$$

For Fig. 2.10*b* with $\phi_2$ circulating in the loop, the coupling changes as follows:

$$-\frac{\lambda}{4!}\phi_1^4 \to -\frac{2\lambda}{4!}\phi_1^2\phi_2^2 \tag{2.333}$$

and as a consequence, the factor $\lambda/2$ in $V$ is replaced by

$$\frac{\lambda}{2} \to \frac{\lambda}{6} \tag{2.334}$$

since the 4! factor is only partially compensated by a combinatoric $(2!)^2$ factor. For Fig. 2.10*c* with the gauge vector in the loop, the coupling is

$$\frac{e^2}{2}\phi_1^2 A_\mu^2 \tag{2.335}$$

and hence, compared to the $\phi_1^4$ expression, one replaces $\lambda$ by $2e^2$ and adds an extra overall factor 3 for the contraction of the propagator numerator.

Adding in the necessary counterterms, we thus arrive at the following expression, including zero-loop and one-loop contributions ($\phi^2 = \phi_1^2 + \phi_2^2$):

$$\begin{aligned} V =& \frac{\lambda}{4!}\phi^4 + A\phi^2 + \frac{1}{4!}B\phi^4 + \left(\frac{4\lambda}{3}+2e^2\right)\frac{\Lambda^2}{64\pi^2}\phi^2 \\ & \frac{\lambda^2\phi^4}{256\pi^2}\left(\ln\frac{\lambda\phi^2}{2\Lambda^2}-\frac{1}{2}\right)+\frac{\lambda^2\phi^4}{230\pi^2}\left(\ln\frac{\lambda\phi^2}{6\Lambda^2}-\frac{1}{2}\right) \\ & + \frac{3e^4\phi^4}{64\pi^2}\left(\ln\frac{e^2\phi^2}{\Lambda^2}-\frac{1}{2}\right) \end{aligned} \tag{2.336}$$

We now impose that the renormalized mass be zero:

$$\left.\frac{\partial^2 V}{\partial\phi^2}\right|_{\phi=0} = 0 \tag{2.337}$$

so that $A$ is chosen to cancel the quadratic terms precisely. The coupling is renormalized at $\phi = M$, some arbitrary mass, such that

$$\left.\frac{\partial^4 V}{\partial\phi^4}\right|_{\phi=M} = \lambda \tag{2.338}$$

Using the identity

$$\frac{d^4}{dx^4}\left(x^4 \ln ax^4\right) = 24 \ln ax^2 + 100 \tag{2.339}$$

one finds that

$$\begin{aligned} B = & \frac{3}{64\pi^2}\left(\frac{10}{9}\lambda^2 + 12e^4\right) - \frac{\lambda^2}{256\pi^2}\phi^2\left(24\ln\frac{\lambda M^2}{2\Lambda^2} + 100\right) \\ & - \frac{\lambda^2}{2304\pi^2}\left(24\ln\frac{\lambda M^2}{6\Lambda^2} + 100\right) - \frac{3e^4}{64\pi^2}\left(24\ln\frac{e^2M^2}{\Lambda^2} + 100\right) \end{aligned} \tag{2.340}$$

Substituting into $V$, one arrives at the final form:

$$v = \frac{\lambda}{4!}\phi^4 + \left(\frac{5}{1152}\frac{\lambda^2}{\pi^2} + \frac{3e^4}{64\pi^2}\right)\phi^4\left(\ln\frac{\phi^2}{M^2} - \frac{25}{6}\right) \tag{2.341}$$

In order that two or more loops be insignificant, we take $\lambda, e \ll 1$. Also, we assume that $\lambda \sim e^4$, so that $\lambda^2 \ll e^4$ and the $\lambda^2$ term may be dropped. Choosing for convenience the renormalization mass to be $M = \langle\phi\rangle$, we have

$$V(\phi) = \frac{\lambda}{24}\phi^4 + \frac{3e^4}{64\pi^2}\phi^4\left(\ln\frac{\phi^2}{\langle\phi\rangle^2} - \frac{25}{6}\right) \tag{2.342}$$

Note that $V(0) = 0$. The derivative gives

$$V'\big(\langle\phi\rangle\big) = \left(\frac{\lambda}{6} - \frac{11}{16}\frac{e^4}{\pi^2}\right)\langle\phi\rangle^3 \tag{2.343}$$

so that if $\langle\phi\rangle \neq 0$, then for $V'(\langle\phi\rangle) = 0$,

$$\lambda = \frac{33}{8}\frac{e^4}{\pi^2} \tag{2.344}$$

With this constraint one finds that

$$V\big(\langle\phi\rangle\big) = -\frac{3}{128}\frac{e^4}{\pi^2}\langle\phi\rangle^4 \tag{2.345}$$

so that the minimum occurs for $\langle\phi\rangle \neq 0$, and there is spontaneous symmetry breaking.

To compute the scalar mass, note that

$$V(\phi) = \frac{3e^4\phi^4}{64\pi^2}\left(\ln\frac{\phi^2}{\langle\phi\rangle^2} - \frac{1}{2}\right) \tag{2.346}$$

$$\frac{\partial V}{\partial\phi} = \frac{3e^4\phi^3}{64\pi^2}\ln\frac{\phi^2}{\langle\phi\rangle^2} \tag{2.347}$$

$$\left.\frac{\partial^2 V}{\partial \phi^2}\right|_{\phi=\langle\phi\rangle} = \frac{3e^4}{8\pi^2}\langle\phi\rangle^2 = M_s^2 \tag{2.348}$$

The vector mass arising from the Higgs mechanism is given by

$$M_V^2 = e^2\langle\phi\rangle^2 \tag{2.349}$$

so that

$$\frac{M_s^2}{M_V^2} = \frac{3}{2\pi}\frac{e^2}{4\pi} \tag{2.350}$$

Thus the one-loop contributions reveal that the theory with massless scalars in the symmetric Lagrangian resembles more closely the $M_0^2 < 0$ case (Higgs) than the $M_0^2 > 0$ case (scalar electrodynamics).

Note, in particular, that one of the two dimensionless parameters $e$ and $\lambda$ has been replaced by a dimensional one, $\langle\phi\rangle$. This phenomenon, where spontaneous breakdown of theory with scale-invariant Lagrangian gives rise to masses, is called *dimensional transmutation*.

**Example 2.** In this example we discuss an $O(N)$ scalar (nongauge) model [37, 38], where $N$ is taken to be very large. This model is one where the tree approximation indicates spontaneous breakdown, and it is only when the one-loop graphs are included that it is discovered that the ground state is, in fact, symmetric; roughly speaking, the situation here is thus the reverse of that for massless scalar electrodynamics. The Lagrangian is ($a = 1, 2, \ldots, N$)

$$L = \frac{1}{2}\partial_\mu\phi^a\partial_\mu\phi^a - \frac{1}{2}M_0^2\phi^a\phi^a - \frac{\lambda_0}{8N}\left(\phi^a\phi^a\right)^2 \tag{2.351}$$

In the tree approximation the potential is

$$V = \frac{1}{2}M_0^2\phi^a\phi^a + \frac{\lambda_0}{8N}\left(\phi^a\phi^a\right)^2 \tag{2.352}$$

We take $\lambda_0 > 0$; otherwise, the spectrum is unbounded below. For $M_0^2 > 0$ the ground state has $\phi^a = 0$. For $M_0^2 < 0$ this potential has a minimum for

$$\langle\phi\rangle^2 = \frac{-2M_0^2 N}{\lambda_0} > 0 \tag{2.353}$$

Putting $\phi^a = \delta_{an}\langle\phi\rangle$ and defining

$$\sigma = \phi_N - \langle\phi\rangle \tag{2.354}$$

$$\pi^a = \phi^a \quad a = 1, 2, 3, \ldots, (N-1) \tag{2.355}$$

one finds that

$$V = \frac{\lambda_0}{8N}\left(\pi^a\pi^a + \sigma^2 + 2\sigma\langle\phi\rangle^2\right) - \frac{M_0^4 N}{2\lambda_0} \tag{2.356}$$

and

$$L = \frac{1}{2}\partial_\mu\pi^a\partial_\mu\pi^a + \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma - V \tag{2.357}$$

Thus in this approximation the spontaneous breaking of $O(N)$ to a residual $O(N-1)$ results in $(N-1)$ Goldstone bosons $\pi^a$. As will become apparent shortly, this approximation is not a good one.

To compute the one-loop corrections, we use a combinatoric trick [37] by adding a term to $L$ as follows:

$$\begin{aligned} L \;=\; & \frac{1}{2}\partial_\mu\phi^a\partial_\mu\phi^a - \frac{1}{2}M_0^2\phi^a\phi^a - \frac{\lambda_0}{8N}\left(\phi^a\phi^a\right)^2 \\ & + \frac{N}{2\lambda_0}\left(x - \frac{1}{2}\frac{\lambda_0}{N}\phi^a\phi^a - M_0^2\right) \end{aligned} \tag{2.358}$$

The new field $x$ is not a dynamical degree of freedom since its equation of motion is simply

$$\frac{\partial L}{\partial x} = 0 \tag{2.359}$$

$$x = \frac{1}{2}\frac{\lambda_0}{N}\phi^a\phi^a + M_0^2 \tag{2.360}$$

The Lagrangian now becomes

$$L = \frac{1}{2}\partial_\mu\phi^a\partial_\mu\phi^a + \frac{N}{2\lambda_0}x^2 - \frac{1}{2}x\phi^a\phi^a - \frac{NM_0^2}{\lambda_0}x - \frac{NM_0^4}{2\lambda_0} \tag{2.361}$$

The only interaction is $-\frac{1}{2}x\phi^a\phi^a$ and the Feynman rules now give $1/N$ factors only for the $x$ propagator. To leading order in $1/N$, therefore, one needs to sum the graphs indicated in Fig. 2.11. This sum gives (in Euclidean space)

$$V(\text{one loop}) = -N\int\frac{d^4k}{(2\pi)^4}\sum\frac{1}{2n}\frac{(x)^n}{(-k^2)^n} \tag{2.362}$$

$$= \frac{N}{2}\int\frac{d^4k}{(2\pi)^4}\ln\left(1 + \frac{x}{k^2}\right) \tag{2.363}$$

$$= \frac{N}{2}\int\frac{d^4k}{(2\pi)^4}\ln\left(k^2 + x\right) \tag{2.364}$$

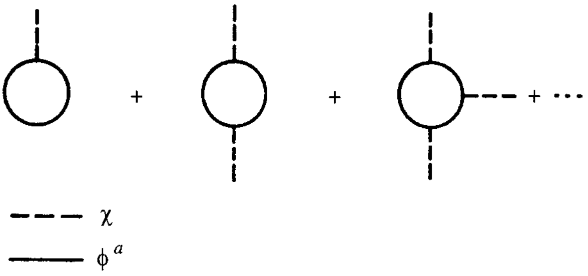



**Figure 2.11** One-loop graphs for $O(N)$ model.

where in the last form we dropped an irrelevant (infinite) constant. The full potential, including the tree diagrams, is

$$V = -\frac{N}{2_0}x^2 + \frac{1}{2}x\phi^a\phi^a + \frac{NM_0^2}{\lambda_0} + \frac{N}{2}\int \frac{d^4k}{(2\pi)^4}\ln\left(k^2 + x\right) \tag{2.365}$$

Now we renormalize $M_0$ and $\lambda_0$ by defining

$$\frac{M_0^2}{\lambda_0} = \frac{M^2}{\lambda} - \frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\frac{1}{k^2} \tag{2.366}$$

$$\frac{1}{\lambda_0} = \frac{1}{\lambda} - \frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\frac{1}{k^2}\frac{1}{k^2 + M^2} \tag{2.367}$$

To express $V$ in terms of these renormalized quantities, we need the integral

$$\int \frac{d^4k}{(2\pi)^4}\left[\ln\left(k^2 + \chi\right) - \frac{\chi}{k^2}\frac{\chi}{2k^2(k^2 + M^2)}\right]$$

$$= \frac{1}{16\pi^2}\int_0^\infty dx\, x\left[\ln(x + \chi) - \frac{\chi}{x} + \frac{\chi}{2x(x + M^2)}\right] \tag{2.368}$$

$$= \frac{1}{16\pi^2}\left[\frac{x^2 - \chi^2}{2}\ln(x + \chi) - \frac{1}{2}x\left(\chi + \frac{1}{2}x\right) + \frac{\chi^2}{2}\ln\left(x + M^2\right)\right]_0^\infty \tag{2.369}$$

$$= \frac{1}{16\pi^2}\left(-\frac{\chi^2}{4} + \frac{\chi^2}{2}\ln\frac{\chi}{M^2}\right) \tag{2.370}$$

where we set $x = k^2$. Using this result, one finds that

$$V = -\frac{N}{2\lambda}\chi^2 + \frac{1}{2}\chi\phi^a\phi^a + \frac{Nm^2}{\lambda_0}\chi + \frac{N\chi^2}{128\pi^2}\left(2\ln\frac{\chi}{M^2} - 1\right) \tag{2.371}$$

At a minimum of $V$,

$$\frac{\partial V}{\partial \chi} = 0 \tag{2.372}$$

and

$$\frac{\partial V}{\partial \phi^a} = 0 \tag{2.373}$$

$$= 2\phi^a \frac{\partial^2 V}{\partial \phi^2} \tag{2.374}$$

so that

$$\phi^a \chi = 0 \tag{2.375}$$

If $\phi^a \neq 0$ at the minimum, $\chi = 0$ and it follows that $V = 0$.

Now we must evaluate $V$ when $\phi^2 = 0$ to check whether the symmetric situation has lower energy. With $\phi^2 = 0$, $\partial V/\partial_\chi = 0$ implies that $\chi = \chi_0$, where

$$\chi_0 - M^2 - \frac{\lambda \chi_0}{32\pi^2} \ln \frac{\chi_0}{M_2} = 0 \tag{2.376}$$

We are interested in the case $m^2 < 0$ and it follows that since $\chi_0$ is real

$$\chi_0 > 0 \tag{2.377}$$

For the potential, we now find that

$$V(\phi^2 = 0, \chi = \chi_0) = -\frac{N}{2\lambda}\chi_0^2 + \frac{Nm^2}{\lambda}\chi_0 + \frac{N\chi_0^2}{128\pi^2}\left(2\ln\frac{\chi_0}{M^2} - 1\right) \tag{2.378}$$

$$= \frac{N\chi_0^2}{128\pi^2}\left(\frac{2m^2}{\chi_0^2 - m^2}\ln\frac{\chi_0}{M^2} - 1\right) \tag{2.379}$$

Now

$$\chi_0^2 - m^2 > 0 \tag{2.380}$$

$$m^2 < 0 \tag{2.381}$$

and hence

$$V\left(\phi^2 = 0\right) < 0 \tag{2.382}$$

Thus the ground state occurs for $\phi^a = 0$ and the $O(N)$ symmetry is not spontaneously broken [38]. In this theory, the one-loop corrections play a crucial role, since in the tree approximation it appeared that there is spontaneous breaking of $O(N)$ down to $O(N-1)$.

**Example 3.** In this example of loop corrections we consider the Glashow–Salam–Weinberg model [39–41]. This model, which we will return to in more detail later, provides the simplest theory that unifies weak and electromagnetic interactions and is renormalizable. The gauge group is $SU(2) \times U(1)$, corresponding to weak isospin and weak hypercharge. There are therefore fore gauge vectors $A_\mu^A$ $(a = 1, 2, 3)$ and $B_\mu$. The left- and right-handed helicity electrons are denoted

$$e_L = \frac{1}{2}(1 - \gamma_5)e \tag{2.383}$$

$$e_R = \frac{1}{2}(1 + \gamma_5)e \tag{2.384}$$

The left-handed component is put together with the neutrino $\nu_L$ in an SU(2) doublet,

$$L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \tag{2.385}$$

with weak hypercharge $y = -1$. The component $R = e_R$ is an SU(2) singlet with $y = -2$. This ensures that the electric charge is always given by

$$Q = t_L^3 + \frac{1}{2}y \tag{2.386}$$

To break the symmetry spontaneously, a complex doublet of Higgs scalars in introduced:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.387}$$

with $y = +1$, and transforming like the left-handed doublet under weak isospin.

The Lagrangian may be written

$$\mathscr{L} = \mathscr{L}_g + \mathscr{L}_l + \mathscr{L}_s + \mathscr{L}_i \tag{2.388}$$

with

$$\mathscr{L}_g = -\frac{1}{4}F_{\mu\nu}^1 F_{\mu\nu}^1 - \frac{1}{4}G_{\mu\nu}G_{\mu\nu} \tag{2.389}$$

$$F_{\mu\nu}^i = \partial_\mu A_\nu^i - \partial_\nu A_\mu^i + g\epsilon^{ijk}A_\mu^j A_\nu^k \tag{2.390}$$

$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.391}$$

$$\mathscr{L}_l = \bar{R}i\gamma_\mu(\partial_\mu + ig'B_\mu)R + \bar{L}i\gamma_\mu\left(\partial_\mu + \frac{i}{2}g'B_\mu - ig\frac{\tau^i}{2}A^i_\mu\right)L \tag{2.392}$$

$$\mathscr{L}_s = \left(\partial_\mu\phi^\dagger + i\frac{g'}{2}B_\mu\phi^\dagger + i\frac{g}{2}\tau^i A^i_\mu\phi^\dagger\right)$$
$$\cdot\left(\partial_\mu\phi - i\frac{g'}{2}B_\mu\phi - i\frac{g}{2}\tau^i A^i_\mu\phi\right) - \mu^2\phi^\dagger\phi - \lambda\left(\phi^\dagger\phi\right)^2 \tag{2.393}$$

$$\mathscr{L}_i = G_e\left(\bar{R}\phi^\dagger L + \bar{L}\phi R\right) \tag{2.394}$$

For the reader unfamiliar with this theory, we first discuss briefly the tree approximation of the Weinberg–Salam model, then proceed to the one-loop corrections, which are our main concern at present.

We may reparametrize the Higgs doublet by

$$\phi = U^{-1}(\xi)\frac{v+\eta}{\sqrt{2}} \tag{2.395}$$

$$U(\xi) = \exp\frac{(-i\boldsymbol{\xi}\cdot\boldsymbol{\tau})}{2v} \tag{2.396}$$

where $v/\sqrt{2}$ is the vacuum value of $\phi$ and the four components of $\phi$ are replaced by $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$. Now we go to the unitary gauge by making the SU(2) gauge transformation specified by $U(\xi)$. That is,

$$\phi' = U\phi \tag{2.397}$$

$$L' = UL \tag{2.398}$$

$$A'_\mu = UA_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1} \tag{2.399}$$

where

$$A_\mu \equiv \mathbf{A}_\mu\cdot\boldsymbol{\tau} \tag{2.400}$$

The fields $B_\mu$ and $R$ remain unchanged. After making this transformation, one finds an electron mass term from

$$\mathscr{L}_i = -G_e\left(\bar{R}\phi^\dagger L + \bar{L}_\phi R\right) \tag{2.401}$$

$$= -\frac{G_e v}{\sqrt{2}}(\bar{e}_R e_L + \bar{e}_L e_R) + \cdots \tag{2.402}$$

$$= -\frac{G_e v}{\sqrt{2}}\bar{e}e + \cdots \tag{2.403}$$

so that

$$G_e = \frac{\sqrt{2}\, M_e}{v} \tag{2.404}$$

and the neutrino remains massless since it has no right-handed counterpart.

The mass of the Higgs scalar $\eta$ is found by using

$$v^2 = \frac{-\mu^2}{\lambda} \tag{2.405}$$

and isolating the quadratic term in $\mathscr{L}_s$, which turns out to be

$$\mathscr{L}_s = \frac{1}{2}\partial_\mu \eta \partial_\mu \eta + \mu^2 \eta^2 + \cdots \tag{2.406}$$

so that

$$m_\eta^2 = -2\mu^2 \tag{2.407}$$

Since $v$ depends only on the ratio $-\mu^2/\lambda$, it appears in the tree approximation that the Higgs mass $m_\eta^2$ can be made arbitrarily large (large $\lambda$). By studying the one-loop contributions, however, we shall arrive at restrictions on this mass.

The masses of the charged intermediate vector bosons $W_\mu^\pm$ defined by

$$W_\mu^\pm = \frac{A_\mu^1 \pm i A_\mu^2}{\sqrt{2}} \tag{2.408}$$

can be found from the quadratic terms in $\mathscr{L}_s$, which are

$$\frac{g^2 v^2}{8}\left[\left(A_\mu^1\right)^2 + \left(A_\mu^2\right)^2\right] = \frac{g^2 v^2}{8}\left(W_\mu^{+*} W_\mu^+ + W_\mu^{-*} W_\mu^-\right] \tag{2.409}$$

so that

$$m_W = \frac{1}{2} g v \tag{2.410}$$

The charged current couplings to the leptons are given by

$$g\bar{L}\gamma \frac{\tau_i}{\mu^2} A_{i\mu} L \quad (i = 1, 2)$$

$$= \frac{g}{\sqrt{2}}\left(\bar{\nu}_L \gamma_\mu e_L W_\mu^+ + \bar{e}_L \gamma_\mu \nu_L W_\mu^-\right) \tag{2.411}$$

$$= \frac{g}{2\sqrt{2}}\left[\bar{\nu}\gamma_\mu (1 - \gamma_5) e W_\mu^+ + \text{h.c.}\right] \tag{2.412}$$

so that in terms of Fermi's constant $G$ we have

$$\frac{g^2}{8m_W^2} = \frac{G}{\sqrt{2}} \tag{2.413}$$

and hence (using $m_W = \frac{1}{2}gv$)

$$v = 2^{-1/4}G^{-1/2} \tag{2.414}$$

Using the value from $\mu$ decay, $G = 1.01 \times 10^{-5}m_p^{-2}$, one finds that

$$v = 248 \text{ GeV} \tag{2.415}$$

This implies that the dimensionless coupling $G_e$ has the value

$$G_e = \frac{\sqrt{2}\,M_e}{v} \tag{2.416}$$

$$\simeq 2 \times 10^{-6} \tag{2.417}$$

Next we turn to the two neutral gauge fields $A_\mu^3$ and $B_\mu$. The photon $A_\mu$ and neutral intermediate vector boson $Z_\mu$ are defined according to

$$B_\mu = \cos\theta_W A_\mu + \sin\theta_W Z_\mu \tag{2.418}$$

$$A_\mu^3 = -\sin\theta_W A_\mu + \cos\theta_W Z_\mu \tag{2.419}$$

where $\theta_W$ is the weak mixing angle, which, as we shall see, satisfies

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \tag{2.420}$$

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \tag{2.421}$$

Picking out the photon couplings from $\mathscr{L}_l$ gives

$$-g'\cos\theta_W \bar{e}_R\gamma_\mu A_\mu e_R - \frac{g'}{2}\cos\theta_W(\bar{\nu}_L\gamma_\mu\nu_L + \bar{e}_L\gamma_\mu e_L)A_\mu$$

$$+\frac{g}{2}\sin\theta_W(\bar{\nu}_L\gamma_\mu\nu_L - \bar{e}_L\gamma_\mu e_L)A_\mu$$

$$= -g'\cos\theta_W A_\mu(\bar{e}_R\gamma_\mu e_R + \bar{e}_L\gamma_\mu e_L) \tag{2.422}$$

$$= -\frac{gg'}{\sqrt{g^2 + g'^2}}A_\mu\bar{e}\gamma_\mu e \tag{2.423}$$

so that

$$e = g \sin\theta_W \tag{2.424}$$

$$= g' \cos\theta_W \tag{2.425}$$

The mass of the neutral intermediate vector boson $Z_\mu$ is provided by the piece of $\mathscr{L}_s$ (the photon is, of course, massless)

$$\frac{v^2}{8}\left(g'B_\mu - gA^3_\mu\right)^2 = \frac{g^2v^2}{8\cos^2\theta_W} Z_\mu Z_\mu \tag{2.426}$$

Substituting for $g$ and $v$ one finds that for the vector masses

$$m_W = \frac{38\ \text{GeV}}{\sin\theta_W} \tag{2.427}$$

$$m_Z = \frac{m_W}{\cos\theta_W} \tag{2.428}$$

$$= \frac{76\ \text{GeV}}{\sin 2\theta_W} \tag{2.429}$$

For example, with the value

$$\theta = 28^\circ\left(\sin^2\theta_W = 0.22\right) \tag{2.430}$$

favored by neutrino scattering experiments, $m_W = 80$ GeV and $m_Z = 91$ GeV. Discovery of charged and neutral vector bosons with these masses at CERN in 1983 provided striking confirmation of the theory.

Finally, the couplings to leptons of the neutral weak current can be read off from $\mathscr{L}_l$. It is found to be

$$-\frac{Z_\mu}{2\sqrt{g^2+g'^2}} \cdot \left[g'^2(2\bar{e}_R\gamma_\mu e_R + \bar{e}_L\gamma_\mu e_L + \bar{\nu}_L\gamma_\mu\nu_L) - g^2(\bar{e}_L\gamma_\mu e_L - \bar{\nu}_L\gamma_\mu\nu_L)\right] \tag{2.431}$$

To summarize the tree approximation, the electromagnetic and weak couplings $e$ and $G$ are not mutually determined but are related through the parameter $\theta_W$, which may be determined empirically. Also there is the arbitrary parameter $m_\eta^2$ for the Higgs scalar.

The one-loop corrections to the effective potential stem from the three types of couplings in

$$-\lambda(\phi^+\phi)^2 + \left(\frac{g'}{2}B_\mu + \frac{g}{2}\tau_i A_{i\mu}\right)^2 \phi^\dagger\phi - G_e\left(\bar{R}\phi^\dagger L + \bar{L}\phi R\right) \tag{2.432}$$

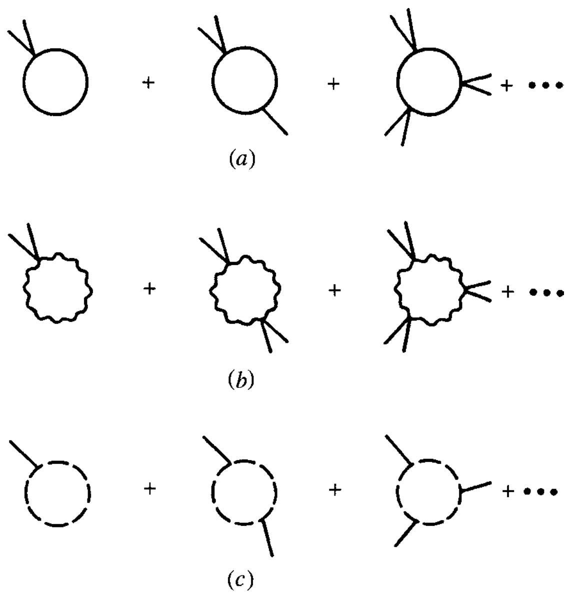



**Figure 2.12** One-loop graphs for Glashow–Salam–Weinberg model: (*a*) scalar loop; (*b*) gauge vector loop; (*c*) fermion loop.

Examples of the corresponding one-loop graphs to be summed are indicated in Fig. 2.12 for (*a*) a scalar loop, (*b*) a gauge-vector loop, and (*c*) a fermion loop. When $\lambda$ is so small that $\lambda \sim e^4$, the relative magnitudes of these three possibilities are $\lambda^2 \sim e^8$, $e^4$, and $G_e^4$, respectively. Since $G_e^4 \sim 16 \times 10^{-24}$, the lepton loop is negligible (even if we replace the electron by a moderately heavy lepton of about 2 GeV). Also, the gauge-vector loop dominates over the scalar loop, so we need compute only the former.

This computation was first performed by Weinberg [42, 43] and is quite similar to our earlier calculation for scalar electrodynamics. There we obtained for the potential

$$V(\phi) = \frac{\lambda}{24}\phi^4 + \frac{3e^4}{64\pi^2}\phi^4\left(\ln\frac{\phi^2}{M^2} - \frac{25}{6}\right) \tag{2.433}$$

which may be rewritten as

$$V(\phi) = \frac{3e^4}{64\pi^2}\phi^4 \ln\frac{\phi^2}{M_f^2} \tag{2.434}$$

where we defined the mass $M_f^2$ by

$$\ln M_f^2 = \ln M^2 + \frac{25}{6} - \frac{8\pi^2\lambda}{9e^4} \tag{2.435}$$

It turns out (see Weinberg, [44] for details) that in the present case of four gauge bosons, one replaces $e^4 = m_v^4/v^4$ by

$$e^4 \to \frac{\sum_v m_v^4}{v^4} \tag{2.436}$$

where $v$ is the vacuum value of the Higgs scalar and $\sum_v m_v^4$ is the trace of the fourth power of the vector mass matrix.

Thus, defining

$$B = \frac{3\sum_v m_v^4}{64\pi^2 v^4} \tag{2.437}$$

one may write

$$V(\phi) = -\frac{1}{2}\mu_R^2\phi^2 + B\phi^4 \ln \frac{\phi^2}{M_f^2} \tag{2.438}$$

where $\mu_R$ is the renormalized scalar mass.

At a stationary point

$$\left.\frac{dV}{d\phi}\right|_{\phi=\langle\phi\rangle} = \mu_R^2\langle\phi\rangle + 4B\langle\phi\rangle^3\left(\ln \frac{\langle\phi\rangle^2}{M_f^2} + \frac{1}{2}\right) \tag{2.439}$$

$$= 0 \tag{2.440}$$

so that either

$$\langle\phi\rangle = 0 \tag{2.441}$$

or

$$\frac{\mu_R^2}{4B} = \langle\phi\rangle^2\left(\ln \frac{\langle\phi\rangle^2}{M_f^2} + \frac{1}{2}\right) \tag{2.442}$$

Differentiating again, one finds (for $\langle\phi\rangle \neq 0$) that

$$\left.\frac{d^2V}{d\phi^2}\right|_{\phi=\langle\phi\rangle} = -4B\langle\phi\rangle^2\left(\ln \frac{\langle\phi\rangle^2}{M_f^2} + \frac{1}{2}\right) + 12B\langle\phi\rangle^2\left(\ln \frac{\langle\phi\rangle^2}{M_f^2} + \frac{7}{6}\right) \tag{2.443}$$

Now, for $\langle\phi\rangle \neq 0$,

$$V(\langle\phi\rangle) = -B\langle\phi\rangle^4\left(\ln \frac{\langle\phi\rangle^2}{M_f^2} + 1\right) \tag{2.444}$$

to be compared with

$$V(0) = 0 \tag{2.445}$$

Thus, in order that the symmetry-breaking vacuum be absolutely stable, one requires that

$$V\big(\langle\phi\rangle\big) < V(0) \tag{2.446}$$

so that

$$\ln\frac{\langle\phi\rangle^2}{M_f^2} > -1 \tag{2.447}$$

Hence the Higgs scalar mass $m_\eta^2$ satisfies

$$m_\eta^2 = \left.\frac{d^2V}{d\phi^2}\right|_{\phi=\langle\phi\rangle} \tag{2.448}$$

$$> 4B\langle\phi\rangle^2 \tag{2.449}$$

$$= \frac{3}{16\pi^2 v^2}\big(2m_W^4 + m_Z^4\big) \tag{2.450}$$

Inserting the expressions for $m_W$, $m_Z$, and $V$ from the tree approximation, one finds that ($\alpha = e^2/4\pi$)

$$v^4 = \big(2G^2\big)^{-1} \tag{2.451}$$

$$m_W^4 = \frac{\pi^2\alpha^2}{2G^2\sin^4\theta_W} \tag{2.452}$$

$$m_Z^4 = \frac{m_W^4}{\cos^4\theta_W} \tag{2.453}$$

and therefore

$$m_\eta^2 > \frac{3\alpha^2}{16\sqrt{2}\,G}\frac{2+\sec^4\theta_W}{\sin^4\theta_W} \tag{2.454}$$

This expression has a minimum for $\theta_W \approx 49.4°$; with this value for $\theta_W$, one finds that

$$m_\eta > 3.75 \text{ GeV} \tag{2.455}$$

For the empirical value $\theta_W = 28°$, one finds that

$$m_\eta > 6.80 \text{ GeV} \tag{2.456}$$

This mass is sufficiently large that Higgs scalars, if they exist, could have escaped detection. (For an early phenomenology of Higgs scalars, see Ref. [45].) To summarize this example, in the tree approximation the Higgs scalar mass may be made arbitrarily small by reducing the self-interaction $\lambda$. But when $\lambda$ is so small that $\lambda \sim e^4$, the gauge vector loop competes with the $\lambda\phi^4$ term and the Higgs mechanism fails unless the Higgs mass exceeds a substantial lower limit.

The discussion so far has made the assumption that the vacuum of the universe is absolutely stable. In 1976 it was shown [46, 47] how to compute the lifetime of a metastable vacuum in quantum field theory and obtain a unique finite answer. The analysis is a four-dimensional extension of the theory of how a three-dimensional bubble forms in a superheated liquid [48–51]. In the latter case a bubble of steam in superheated water has a negative volume energy because steam is the lower energy phase above boiling point but a positive surface energy because of surface tension, and hence there is a critical minimum radius to be reached before boiling is precipitated. For a four-dimensional hypersphere the relevant bubble action is

$$A = -\frac{1}{2}\pi^2 R^4 \epsilon + 2\pi^2 R^3 S_1 \tag{2.457}$$

where $\epsilon$ and $S_1$ are the volume and surface energy densities, respectively. The stationary value of this action is given by

$$A_m = \frac{27}{2}\frac{\pi^2 S_1^4}{\epsilon^3} \tag{2.458}$$

corresponding to a critical radius $R_m = 3S_1/\epsilon$. The energy density $\epsilon$ is just the difference between the effective potential evaluated for the two vacua. The action per unit hypersurface $S_1$ is given by the action for tunneling through the potential barrier $V(\phi)$. In imaginary time this is the classical action for motion in a potential $-V(\phi)$; by considering an infinitesimal change in $R$ it is easy to see that $S_1$ is also the hypersurface energy per unit hyperarea.

The number of such bubbles to form in the past can be estimated as

$$N = \left(V_u \Delta^4\right) \exp(-A_m) \tag{2.459}$$

where $V_u$ is the space-time volume of the backward light cone ($V_u \sim 10^{112}$ cm$^4$) and $\Delta$ is the scale characterizing the bubble; we may take $\Delta \sim R_m^{-1}$.

In this way, one may require that $N \ll 1$ be a condition for a nonglobal minimum of $V(\phi)$ to be a candidate for the vacuum. This then has numerous consequences. For example, in the early universe this idea plays a role in inflation [52] (see also Refs. [53] and [54]). For the electroweak theory, it allows in principle a lower value for the Higgs mass than given in Refs. [42] and [43].

## 2.6 Summary

We have arrived at the Feynman rules for pure Yang–Mills theory in the covariant gauges, using the path integral method and making the ansatz that the volume of the gauge group orbits in the manifold of gauge fields could be divided out as a field-independent multiplicative factor. For the case of the Coulomb gauge, this recipe could be compared successfully to the canonical approach.

The result may be summarized by an effective Lagrangian containing two new terms: One is the gauge-fixing term and the other is the ghost term. The ghosts are fictitious scalars satisfying anticommutation relations and appearing only in closed loops wherever a closed vector loop also occurs.

Adding one-loop corrections, including in the standard model, can alter the nature of the spontaneous symmetry breaking.

## References


**1** P.A.M. Dirac, *Phys. Z. Soviet Union* 3, 64 (1933).
**2** R.P. Feynman, *Rev. Mod. Phys.* 20, 37 (1948).
**3** R.P. Feynman, A.R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw–Hill, New York, 1965.
**4** I. Schwinger, ed., *Selected Papers in Quantum Electrodynamics*, Dover, New York, 1958.
**5** E. Rutherford, *Philos. Mag.* 21, 669 (1911).
**6** G.C. Wick, *Phys. Rev.* 80, 268 (1950).
**7** R.P. Feynman, *Acta Phys. Pol.* 24, 697 (1963).
**8** B.S. DeWitt, *Phys. Rev.* 160, 1113 (1967).
**9** B.S. DeWitt, *Phys. Rev.* 162, 1195 (1967).
**10** B.S. DeWitt, *Phys. Rev.* 162, 1239 (1967).
**11** S. Mandelstam, *Ann. Phys.* 19, 1, 25 (1962).
**12** S. Mandelstam, *Phys. Rev.* 175, 1580 (1968).
**13** S. Mandelstam, *Phys. Rev.* 175, 1604 (1968).
**14** C.N. Yang, R.L. Mills, *Phys. Rev.* 96, 191 (1954).
**15** I.B. Khriplovich, *Sov. J. Nucl. Phys.* 10, 235 (1970).
**16** V.N. Gribov, *Nucl. Phys.* B139, 1 (1978).
**17** R.D. Peccei, *Phys. Rev. D* 17, 1097 (1978).
**18** L.D. Faddeev and V.N. Popov, *Phys. Lett.* B25, 29 (1967).
**19** L.D. Faddeev, *Theor. Math. Phys.* 1, 3 (1969).
**20** L.D. Faddeev, in: *Methods in Field Theory*, R. Balian and J. Zinn-Justin, eds., North–Holland, Amsterdam, 1976, p. 1.
**21** H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Phys. Lett.* B47, 365 (1973).
**22** D.J. Gross and F. Wilczek, *Phys. Rev.* D8, 3497 (1973).
**23** S. Weinberg, *Phys. Rev. Lett.* 31, 494 (1973).
**24** E.S. Fradkin and I.V. Tyutin, *Phys. Lett.* B30, 562 (1969).
**25** E.S. Fradkin and I.V. Tyutin, *Phys. Rev.* D2, 2841 (1970).
**26** C.M. Bender, T. Eguchi, and H. Pagels, *Phys. Rev.* D17, 1086 (1978).
**27** M.B. Halpern and J. Koplik, *Nucl. Phys.* B132, 239 (1978).
**28** R.L. Arnowitt and S.I. Fickler, *Phys. Rev.* 127, 1821 (1962).
**29** F.J. Dyson, *Phys. Rev.* 75, 1736 (1949).
**30** N.N. Bogoliubov and D.V. Shirkov, *Theory of Quantized Fields*, Wiley–Interscience, New York, 1959, Chap. 4.
**31** S.S. Schweber, (1961) *Relativistic Quantum Field Theory*, Harper & Row, New York, Chap. 16.
**32** G. 't Hooft, *Nucl. Phys.* B33, 173 (1971).
**33** G. 't Hooft, *Nucl. Phys.* B35, 167 (1971).
**34** Y. Nambu, *Phys. Lett.* B26, 626 (1968).
**35** S. Coleman, J. Wess, and B. Zumino, *Phys. Rev.* 177, 2239 (1969).
**36** S. Coleman and E. Weinberg, *Phys. Rev.* D7, 1888 (1973).
**37** S. Coleman, R. Jackiw, and H.D. Politzer, *Phys. Rev.*, D10, 2491 (1974).
**38** R.W. Haymaker, *Phys. Rev.* D13, 968 (1976).
**39** S.L. Glashow, *Nucl. Phys.* 22, 579 (1961).
**40** S. Weinberg, *Phys. Rev. Lett.* 19, 1264 (1967).
**41** A. Salam, in: *Elementary Particle Theory*, N. Svartholm, ed., Almqvist & Wiksell, Stockholm, 1968.

**42** S. Weinberg, *Phys. Rev. Lett.* 36, 294 (1976).

**43** A.D. Linde, *JETP Lett.* 23, 73 (1976).

**44** S. Weinberg, *Phys. Rev.* D7, 2887 (1973).

**45** J. Ellis, M.K. Gaillard, and D.V. Nanopoulos, *Nucl. Phys.* B106, 292 (1976).

**46** P.H. Frampton, *Phys. Rev. Lett.* 37, 1378 (1976).

**47** P.H. Frampton, *Phys. Rev.* D15, 2922 (1977).

**48** N.G. Van Kampen, *Phys. Rev.* 135, A362 (1964).

**49** J.S. Langer, *Ann. Phys.* 41, 108 (1967).

**50** T.D. Lee and G.C. Wick, *Phys. Rev.* D9, 2291 (1974).

**51** V.B. Voloshin, L.Yu. Kobzarev, and L.B. Okun, *Sov. J. Nucl. Phys.* 20, 644 (1975).

**52** A.H. Guth, *Phys. Rev.* D23, 347 (1981).

**53** D. Kazanas, *Ap. J.* 241, L59 (1980).

**54** K. Sato, *Monogr. Nat. R. Astron. Soc.* 195, 467 (1981).

# 3
# Renormalization

## 3.1
## Introduction

Here we discuss the renormalization of non-Abelian gauge theories with both exact symmetry and spontaneous symmetry breaking. First, the method of dimensional regularization is described and applied to several explicit examples. The existence of a regularization technique that maintains gauge invariance is crucial in what follows. The presence of triangle anomalies can destroy gauge invariance and hence renormalizability. This is explained together with the general method to cancel the anomalies in an $\mathrm{SU}(2) \times \mathrm{U}(1)$ theory.

The introduction of the Becchi–Rouet–Stora (BRS) form of the local gauge transformation greatly facilitates the study of Green's functions, and this use of anticommuting $c$-numbers and of the Faddeev–Popov ghost field as the gauge function is explained. Using this, the iterative proof of renormalizability is given to all orders, first for pure Yang–Mills, then with fermion and scalar matter fields. The treatment is not totally rigorous (in particular, the details of handling subdivergences are omitted), but all the principal steps are given. Finally, for massive vectors arising from spontaneous breaking, the demonstration of renormalizability is given using the 't Hooft gauges.

## 3.2
## Dimensional Regularization

When we make perturbative calculations in a gauge field theory, there are loop diagrams that involve ultraviolet divergent momentum integrals. The renormalization program separates out these divergent parts and reinterprets them as multiplicative renormalizations of the fields, couplings, and masses in the original (bare) Lagrangian. The bare Lagrangian is written as a sum of the renormalized Lagrangian plus counterterms

$$\mathscr{L}_B(\phi_B) = \mathscr{L}_R(\phi_R) + \text{counterterms} \tag{3.1}$$

The renormalized Lagrangian is of the same functional form as the bare Lagrangian, with the bare couplings and masses replaced by renormalized ones:

$$\mathscr{L}_R(\phi_R, g_R, M_R) = \mathscr{L}_B(\phi_B, g_B, M_B) \tag{3.2}$$

Thus the number of types of divergence that occur must not exceed the (finite) number of terms in the Lagrangian.

The calculational procedure will now be to use only the renormalized Lagrangian $\mathscr{L}_R$, with canonical vertices and propagators, and to omit the divergent parts, which are already accounted for in the renormalization stage.

A very important point specific to a gauge theory is that $\mathscr{L}_R$ should itself be invariant under a set of local gauge transformations isomorphic to those that leave $\mathscr{L}_B$ invariant. This is because we insist on gauge invariance order by order in the loop expansion, and the renormalization involves an infinite reordering of this expansion.

While separating out the divergent parts, it is necessary to modify the integrals so that they are finite and can be manipulated in equations. This modification or *regularization* of the integrals will eventually be removed, in conjunction with the renormalization process, before comparison with experiment. The regularization must violate some physical laws; otherwise, the regularized theory would be satisfactory and finite without renormalization. In practice, one likes to maintain Lorentz invariance (except on the lattice) and, for the reason given in the previous paragraph, gauge invariance. For quantum electrodynamics, the Pauli–Villars method [1] (see also Refs. [2–5]) is adequate. Where there is a non-Abelian gauge group, however, the Pauli–Villars technique proves, in general, inadequate to maintain gauge invariance. Instead, a better method is that of dimensional regularization where the space-time dimensionality is continued analytically from the real physical value, four, to a complex generic value, $n$. With real $n$ sufficiently small, the ultraviolet behavior is convergent; eventually, the limit $n \to 4$ is taken, with the poles in $(n-4)^{-1}$ defining the counterterms. The Ward identities implied by gauge invariance are maintained for general $n$. (We assume that closed fermion loops involving $\gamma_5$ are absent; this special problem will be mentioned again later.) The dimensional regularization procedure was introduced by several people independently [6–14] (see also related earlier work [15, 16]).

Before applying dimensional regularization, one uses the Feynman parameter formulas [17] to rewrite the propagators in a more convenient representation. The simplest such formula is

$$\frac{1}{ab} = \int_0^1 \frac{dx}{[ax + b(1-x)]^2} \tag{3.3}$$

This is verified at once by rewriting the right-hand side as

$$\int_1^\infty \frac{dy}{[a+b(y-1)]^2} = \int_0^\infty \frac{dz}{(a+bz)^2} \tag{3.4}$$

$$= \left.\frac{-1}{b(a+bz)}\right|_0^\infty \tag{3.5}$$

$$= \frac{1}{ab} \tag{3.6}$$

where $y = 1/x$ and $z = y - 1$.

Taking the derivative of Eq. (3.3) with respect to $a$ gives

$$\frac{1}{a^2 b} = 2\int_0^1 \frac{x\,dx}{[ax+b(1-x)]^3} \tag{3.7}$$

More generally, we may take the derivative

$$\frac{d^{\alpha-1}}{da^{\alpha-1}}\frac{d^{\beta-1}}{db^{\beta-1}} \tag{3.8}$$

acting on Eq. (3.3) to find

$$\frac{(\alpha-1)!(\beta-1)!}{a^\alpha b^\beta} = (\alpha+\beta-1)! \cdot \int_0^1 dx \frac{x^{\alpha-1}(1-x)^{\beta-1}}{[ax+b(1-x)]^{\alpha+\beta}} \tag{3.9}$$

or, equivalently,

$$\frac{1}{a^\alpha b^\beta} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 dx \frac{x^{\alpha-1}(1-x)^{\beta-1}}{[ax+b(1-x)]^{\alpha+\beta}} \tag{3.10}$$

Following similar steps, we may arrive at the following generalizations of Eqs. (3.3) and (3.10):

$$\frac{1}{abc} = 2\int_0^1 dx \int_0^{1-x} dy \frac{1}{[a(1-x-y)+bx+cy]^3} \tag{3.11}$$

and

$$\frac{1}{a^\alpha b^\beta c^\gamma} = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} \cdot \int_0^1 dx \int_0^{1-x} dy \frac{(1-x-y)^{\alpha-1}x^{\beta-1}y^{\gamma-1}}{[a(1-x-y)+bx+cy]^{\alpha+\beta+\gamma}} \tag{3.12}$$

respectively.

As the most general case of the Feynman parameter formulas we may similarly derive

$$\frac{1}{a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_n^{\alpha_n}} = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_n)}$$
$$\cdot \int_0^1 dx_1 \int_0^{1-x_1} dx_2 \cdots \int_0^{1-x_1-x_2\cdots-x_{n-1}} dx_{n-1}$$
$$\cdot \frac{(1 - x_1 - x_2 - \cdots - x_{n-1})^{\alpha_1 - 1} x_1^{\alpha_2 - 1} \cdots x_{n-1}^{\alpha_n - 1}}{[a_1(1 - x_1 - \cdots - x_{n-1}) + a_2 x_1 + \cdots + a_n x_{n-1}]^{\Sigma\alpha_1}} \tag{3.13}$$

We are ready now to consider our first integral in a generic dimension. The following formula is important to establish

$$\int \frac{d^n k}{(k^2 + 2k \cdot Q - M^2)^\alpha} = \frac{i\pi^{n/2}\Gamma(\alpha - n/2)}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}} \tag{3.14}$$

The $n$ dimensions comprise one time and $(n-1)$ space dimensions, so let us write the $n$-vector $k_\mu = (k_0, \mathbf{w})$, whereupon

$$\int d^n k \equiv \int_{-\infty}^{\infty} dk_0 \int_0^{\infty} w^{n-2} dw \int_0^{2\pi} d\theta_1 \int_0^{\pi} d\theta_2 \sin\theta_2 \int_0^{\pi} d\theta_3 \sin^2\theta_3$$
$$\cdots \int_0^{\pi} d\theta_{n-2} \sin^{n-3}\theta_{n-2} \tag{3.15}$$

Let us evaluate the integral in the rest frame $Q_\mu = (\mu, \mathbf{0})$. Then use

$$\int \sin^m\theta \, d\theta = \frac{\sqrt{\pi}\Gamma(\frac{m}{2} + \frac{1}{2})}{\Gamma(\frac{m}{2} + 1)} \tag{3.16}$$

to arrive at

$$I_n(Q) = \int \frac{d^n k}{(k^2 + 2k \cdot Q - M^2)^\alpha} \tag{3.17}$$

$$= \frac{2\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2})} \int_{-\infty}^{\infty} dk_0 \int_0^{\infty} \frac{w^{n-2}\, dw}{(k_0^2 - w^2 + 2\mu k_0 - M^2)^\alpha} \tag{3.18}$$

Here we used $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and hence

$$2(\sqrt{\pi})^{n-2} \frac{\Gamma(\frac{1}{2})}{\Gamma(1)} \frac{\Gamma(1)}{\Gamma(\frac{3}{2})} \cdots \frac{\Gamma(\frac{n-3}{2} + \frac{1}{2})}{\Gamma(\frac{n-3}{2} + 1)} = \frac{2\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2})} \tag{3.19}$$

Now change the variable to $k_\mu = (k + Q)_\mu$, giving

$$I_n(Q) = \frac{2\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2})} \int_{-\infty}^{\infty} dk_0 \int_0^{\infty} \frac{w^{n-2}\, dw}{[k_0^2 - w^2 + (Q^2 + M^2)]^\alpha} \tag{3.20}$$

It is useful to recall a representation for the Euler $B$ function (see, e.g., Ref. [18]):

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{3.21}$$

$$= 2\int_0^\infty t^{2x-1}\left(1+t^2\right)^{-x-y} dx \tag{3.22}$$

with the substitutions

$$x = \frac{1}{2}(1+\beta) \tag{3.23}$$

$$y = \alpha - \frac{1}{2}(1+\beta) \tag{3.24}$$

$$t = \frac{\xi}{M} \tag{3.25}$$

one has

$$\int_0^\infty d\xi \frac{\xi^\beta}{(\xi^2+M^2)^\alpha} = \frac{1}{2}\frac{\Gamma(\frac{1+\beta}{2})\Gamma(\alpha-\frac{1+\beta}{2})}{\Gamma(\alpha)(M^2)^{\alpha-(1/2)(1+\beta)}} \tag{3.26}$$

Employing Eq. (3.26) in Eq. (3.20), one arrives at

$$\begin{aligned} I_n(Q) = & \frac{2(-1)^{-\alpha}\pi^{(n-1)/2}\Gamma(\alpha-\frac{1}{2}(n-1))}{\Gamma(\alpha)} \\ & \cdot \int_0^\infty dk_0 \frac{1}{[Q^2+M^2-k_0^2]^{\alpha-(1/2)(n-1)}} \end{aligned} \tag{3.27}$$

The $k_0$ integration can easily be done, using Eq. (3.26) again, to obtain, finally,

$$I_n(Q) = \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2-M^2)^{\alpha-(n/2)}}\left[\Gamma\left(\alpha-\frac{n}{2}\right)\right] \tag{3.28}$$

which is the required result. This formula can equivalently be written

$$I_n(Q) = \frac{i(-1)^{-\alpha}(-\pi)^{n/2}}{\Gamma(\alpha)(Q^2+M^2)^{\alpha-(n/2)}}\left[\Gamma\left(\alpha-\frac{n}{2}\right)\right] \tag{3.29}$$

In this alternative form, the term in brackets [for the derivative formulas (3.30) through (3.40)] would be rewritten in terms of $(Q^2+M^2)$ rather than $(-Q^2-M^2)$; we mention this alternative because both forms appear in the literature.

By differentiating the result (3.28), we obtain successively a number of valuable identities:

$$\int \frac{d^n k\, k_\mu}{(k^2 + 2k\cdot Q - M^2)^\alpha} = \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}}\left[-Q_\mu \Gamma\left(\alpha - \frac{n}{2}\right)\right] \tag{3.30}$$

$$\begin{aligned}\int \frac{d^n k\, k_\mu k_\nu}{(k^2 + 2k\cdot Q - M^2)^\alpha} &= \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}}\left[Q_\mu Q_\nu \Gamma\left(\alpha - \frac{n}{2}\right)\right.\\ &\quad \left. + \frac{1}{2} g_{\mu\nu}\left(-Q^2 - M^2\right)\Gamma\left(\alpha - 1 - \frac{n}{2}\right)\right]\end{aligned} \tag{3.31}$$

$$\begin{aligned}\int \frac{d^n k\, k_\mu k_\nu k_\kappa}{(k^2 + 2k\cdot Q - M^2)^\alpha} &= \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}}\left[Q_\mu Q_\nu Q_\kappa \Gamma\left(\alpha - \frac{n}{2}\right)\right.\\ &\quad - \frac{1}{2}(g_{\mu\nu} Q_\kappa + g_{\nu\kappa} Q_\mu + g_{\kappa\mu} Q_\nu)\left(-Q^2 - M^2\right)\\ &\quad \left.\cdot\,\Gamma\left(\alpha - 1 - \frac{n}{2}\right)\right]\end{aligned} \tag{3.32}$$

$$\begin{aligned}\int \frac{d^n k\, k_\mu k_\nu k_\kappa k_\lambda}{(k^2 + 2k\cdot Q - M^2)^\alpha} &= \frac{i\pi^{n/2}}{\Gamma(\alpha)(Q^2 - M^2)^{\alpha - n/2}}\left[Q_\mu Q_\nu Q_\kappa Q_\lambda \Gamma\left(\alpha - \frac{n}{2}\right)\right.\\ &\quad + \frac{1}{2}(Q_\mu Q_\nu g_{\kappa\lambda} + \text{permutations})\left(-Q^2 - M^2\right)\\ &\quad \cdot\,\Gamma\left(\alpha - 1 - \frac{n}{2}\right)\\ &\quad + \frac{1}{4}(g_{\mu\nu} g_{\kappa\lambda} + \text{permutations})\left(-Q^2 - M^2\right)^2\\ &\quad \left.\cdot\,\Gamma\left(\alpha - 2 - \frac{n}{2}\right)\right]\end{aligned} \tag{3.33}$$

One can, of course, take contractions of the last three formulas to find (recall that $g_{\mu\nu} g^{\mu\nu} = \delta^\mu_\mu = n$)

$$\begin{aligned}\int d^n k \frac{k^2}{(k^2 + 2k\cdot Q - M^2)^\alpha} &= \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}}\left[Q^2 \Gamma\left(\alpha - \frac{n}{2}\right)\right.\\ &\quad \left. + \frac{n}{2}\left(-Q^2 - M^2\right)\Gamma\left(\alpha - 1 - \frac{n}{2}\right)\right]\end{aligned} \tag{3.34}$$

$$\int d^n k \frac{k^2 k_\mu}{(k^2 + 2k \cdot Q - M^2)^\alpha} = \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}} \cdot \left[ -Q^2 Q_\mu \Gamma\left(\alpha - \frac{n}{2}\right) + \frac{1}{2}(n+2) Q_\mu \left(-Q^2 - M^2\right) \cdot \Gamma\left(\alpha - 1 - \frac{n}{2}\right) \right] \tag{3.35}$$

$$\int d^n k \frac{k^2 k_\mu k_\nu}{(k^2 + 2k \cdot Q - M^2)^\alpha} = \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}} \cdot \left[ Q^2 Q_\mu Q_\nu \Gamma\left(\alpha - \frac{n}{2}\right) + \frac{1}{2}\left(Q^2 g_{\mu\nu} + (n+4) Q_\mu Q_\nu\right)\left(-Q^2 - M^2\right) \cdot \Gamma\left(\alpha - 1 - \frac{n}{2}\right) + \frac{1}{4}(n+2)\left(-Q^2 - M^2\right)^2 g_{\mu\nu} \cdot \Gamma\left(\alpha - 2 - \frac{n}{2}\right) \right] \tag{3.36}$$

$$\int d^n k \frac{k^4}{(k^2 + 2k \cdot Q - M^2)^\alpha} = \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^{\alpha - n/2}} \cdot \left[ -Q^4 \Gamma\left(\alpha - \frac{n}{2}\right) + (n+2) Q^2 \left(-Q^2 - M^2\right) \Gamma\left(\alpha - 1 - \frac{n}{2}\right) + \frac{1}{4} n (n+2) \left(-Q^2 - M^2\right)^2 \cdot \Gamma\left(\alpha - 2 - \frac{n}{2}\right) \right] \tag{3.37}$$

A general formula is not difficult to obtain, and we quote the result below. For general $p$ one has

$$\int d^n k \frac{k_{\mu_1} k_{\mu_2} \cdots k_{\mu_p}}{(k^2 + 2k \cdot Q - M^2)^{\alpha - n/2}} = \frac{i\pi^{n/2}}{\Gamma(\alpha)(-Q^2 - M^2)^\alpha} T^{(p)}_{\mu_1 \mu_2 \cdots \mu_p} \tag{3.38}$$

where the tensor $T^{(p)}$ is given by

$$
\begin{aligned}
T^{(p)}_{\mu_1\mu_2\cdots\mu_p} = (-1)^p \Bigg[ & Q_{\mu_1} Q_{\mu_2} \cdots Q_{\mu_p} \Gamma\left(\alpha - \frac{n}{2}\right) \\
& + \frac{1}{2} \sum_{\substack{\text{permutations} \\ \{\mu_i\}}} (g_{\mu_1\mu_2} Q_{\mu_3} \cdots Q_{\mu_p}) \left(-Q^2 - M^2\right) \\
& \cdot \Gamma\left(\alpha - 1 - \frac{n}{2}\right) \\
& + \frac{1}{4} \sum_{\substack{\text{permutations} \\ \{\mu_i\}}} (g_{\mu_1\mu_2} g_{\mu_3\mu_4} Q_{\mu_5} \cdots Q_{\mu_p}) \left(-Q^2 - M^2\right)^2 \\
& \cdot \Gamma\left(\alpha - 2 - \frac{n}{2}\right) \\
& + \cdots + \left(\frac{1}{2}\right)^{p/2} \sum_{\substack{\text{permutations} \\ \{\mu_i\}}} (g_{\mu_1\mu_2} g_{\mu_3\mu_4} \cdots g_{\mu_{p-1}\mu_p}) \\
& \cdot \left(-Q^2 - M^2\right)^{p/2} \Gamma\left(\alpha - \frac{p}{2} - \frac{n}{2}\right) \Bigg]
\end{aligned}
\tag{3.39}
$$

Equation (3.39) has been written for $p$ even; the only change for $p$ odd is that the last term in brackets should be written

$$
\begin{aligned}
& + \frac{1^{[p/2]}}{2} \sum_{\substack{\text{permutations} \\ \{\mu_i\}}} (g_{\mu_1\mu_2} \cdots g_{\mu_{p-2}\mu_{p-1}} Q_{\mu_p}) \left(-Q^2 - M^2\right)^{[p/2]} \\
& \cdot \Gamma\left(\alpha - \left[\frac{p}{2}\right] - \frac{n}{2}\right)
\end{aligned}
\tag{3.40}
$$

where $[x]$ means the largest integer not greater than $x$.

Equation (3.38) can be derived by induction on $p$.

This completes the principal artillery we need to address the computation of Feynman diagrams in gauge theory, Eq. (3.13), and the regularization formula (3.38). The Feynman rules for pure Yang–Mills theory were provided in Chapter 2; with matter fields, the propagators are given by Bjorken and Drell [19, App. B], whose conventions we use, and the vertices are deduced simply from the Lagrangian.

To illustrate how dimensional regularization is used, we shall work through four examples in some detail. The first three are lowest (fourth)-order divergences in quantum electrodynamics and are treated using Pauli–Villars regularization in the classic text of Bjorken and Drell [19, Chap. 8]. The reader should become convinced that the dimensional method is superior and that Chapter 8 of Ref. [19] could have been shorter had the authors known of dimensional regularization, which was invented about eight years after Ref. [19] was published.

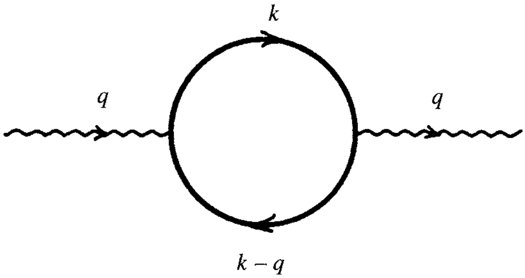



**Figure 3.1** Vacuum polarization in QED.

**Example 1.** The most severe divergence of QED is the quadratic one associated with the vacuum polarization diagram of Fig. 3.1. This diagram modifies the bare photon propagator according to

$$\frac{-ig_{\mu\nu}}{q^2+i\epsilon} \to \frac{-i}{q^2-i\epsilon} I_{\mu\nu}(q) \frac{-i}{q^2+i\epsilon} \tag{3.41}$$

where

$$I_{\mu\nu}(q) = (-1)\int \frac{d^4k}{(2\pi)^2} \mathrm{Tr}\Bigg[(-ie\gamma_\mu)\frac{i}{\not{k}-m+i\epsilon}$$

$$\cdot(-ie\gamma_\nu)\frac{i}{(\not{k}-\not{q})-m+i\epsilon}\Bigg] \tag{3.42}$$

$$= \frac{-e^2}{(2\pi)^2}\int \frac{d^4k\,\mathrm{Tr}[\gamma_\mu(\not{k}+m)\gamma_\nu(\not{k}-\not{q}+m)]}{(k^2-m^2+i\epsilon)[(k-q)^2-m^2+i\epsilon]} \tag{3.43}$$

We will calculate $I_{\mu\nu}(q)$, including the finite $q^2$ dependent Uehling term [20], by two methods. The first uses Pauli–Villars regularization (following Ref. [19]); the second employs dimensional regularization.

The integral in Eq. (3.43) diverges as $\int^\infty(dk\,k)$ (i.e., quadratically). We first evaluate the Dirac trace

$$4T_{\mu\nu} = \mathrm{Tr}\big[\gamma_\mu(\not{k}+m)\gamma_\nu(\not{k}-\not{q}+m)\big] \tag{3.44}$$

$$T_{\mu\nu} = k_\mu(k-q)_\nu + k_\nu(k-q)_\mu - g_{\mu\nu}\big(k^2-k\cdot q-m^2\big) \tag{3.45}$$

Next we use

$$\frac{1}{k^2-m^2+i\epsilon} = -i\int_0^\infty dz\exp\big[iz\big(k^2-m^2+i\epsilon\big)\big] \tag{3.46}$$

to arrive at

$$I_{\mu\nu}(q) = 4e \int_0^\infty dz_1\, dz_2 \int \frac{d^4k}{(2\pi)^4} T_{\mu\nu}(k, q) \cdot \exp\left[iz_1\left(k^2 - m^2 + i\epsilon\right) + iz_2\left((k-q)^2 - m^2 + i\epsilon\right)\right] \tag{3.47}$$

Now complete the square in the exponent by changing the variable to

$$l_\mu = k_\mu - \frac{z_2 q_\mu}{z_1 + z_2} \tag{3.48}$$

so that $d^4k = d^4l$ and

$$I_{\mu\nu}(q) = 4e^2 \int_0^\infty dz_1\, dz_2 \int \frac{d^4k}{(2\pi)^4} T_{\mu\nu}(l, q) e^{i(z_1+z_2)l^2} \cdot \exp\left\{i\left[\frac{q^2 z_1 z_2}{z_1 + z_2} - \left(m^2 - i\epsilon\right)(z_1 + z_2)\right]\right\} \tag{3.49}$$

with

$$T_{\mu\nu}(l, q) = 2l_\mu l_\nu + \frac{z_2 - z_1}{z_2 + z_1}(q_\mu l_\nu + q_\nu l_\mu) - \frac{2z_1 z_2}{(z_1 + z_2)^2} q_\mu q_\nu - g_{\mu\nu}\left[l^2 + \frac{z_2 - z_1}{z_2 + z_1} l \cdot q - \frac{z_1 z_2}{(z_1 + z_2)^2} q^2 - m^2\right] \tag{3.50}$$

Now do the $l$ integration using

$$\int \frac{d^4l}{(2\pi)^4}[1, l_\mu, l_\mu l_\nu] e^{il^2(z_1+z_2)} = \frac{1}{16\pi^2 i(z_1 + z_2)^2}\left[1, 0, \frac{ig_{\mu\nu}}{2(z_1 + z_2)}\right] \tag{3.51}$$

to arrive at ($\alpha = e^2/4\pi$)

$$I_{\mu\nu}(q, m) = -\frac{i\alpha}{\pi} \int_0^\infty \frac{dz_1\, dz_2}{(z_1 + z_2)} \cdot \exp\left\{i\left[\frac{q^2 z_1 z_2}{z_1 + z_2} - \left(m^2 - i\epsilon\right)(z_1 + z_2)\right]\right\} \cdot \left\{2\left(g_{\mu\nu} q^2 - q_\mu q_\nu\right)\frac{z_1 z_2}{(z_1 + z_2)^2} + g_{\mu\nu}\left[\frac{-i}{z_1 + z_2} - \frac{q^2 z_1 z_2}{(z_1 + z_2)^2} + m^2\right]\right\} \tag{3.52}$$

This integral is (of course) still divergent, so we now use the Pauli–Villars method, which is to introduce fictitious heavy electrons of masses $M_i$ ($i = 2, 3, \ldots$; here $M_1 = m_e$). We then consider

$$\bar{I}_{\mu\nu}(q) = \sum_i c_i I_{\mu\nu}(q, M_i) \tag{3.53}$$

The coefficients $C_i$ are chosen such that the integral converges, and the fictitious regulator masses will eventually be allowed to become arbitrarily large.

The first task is to show that the non-gauge-invariant term in Eq. (3.52) vanishes. The coefficient of $g_{\mu\nu}$ is

$$\sum_i c_i \int \frac{dz_1\,dz_2}{(z_1+z_2)^2} \exp\left\{i\left[\frac{q^2 z_1 z_2}{z_1+z_2} - \left(M_i^2 - i\epsilon\right)(z_1+z_2)\right]\right\} \cdot \left[M_i^2 - \frac{i}{z_1+z_2} - \frac{q^2 z_1 z_2}{(z_1+z_2)^2}\right] \tag{3.54}$$

We change variables $z_i \to \lambda z_i$ to rewrite Eq. (3.54) as

$$\sum_i c_i \int_0^\infty \frac{dz_1\,dz_2}{(z_1+z_2)^2}\left[M_i^2 - \frac{i}{\lambda(z_1+z_2)} - \frac{q^2 z_1 z_2}{(z_1+z_2)^2}\right] \cdot \exp\left\{i\lambda\left[\frac{q^2 z_1 z_2}{z_1+z_2} - \left(M_i^2 - i\epsilon\right)(z_1+z_2)\right]\right\} \tag{3.55}$$

$$= i\lambda\frac{\partial}{\partial\lambda}\int_0^\infty \frac{dz_1\,dz_2}{\lambda(z_1+z_2)^3} \cdot \sum_i c_i \exp\left\{i\lambda\left[\frac{q^2 z_1 z_2}{z_1+z_2} - \left(M_i^2 - i\epsilon\right)(z_1+z_2)\right]\right\} \tag{3.56}$$

Finally, rescale $\lambda z_i \to z_i$ again to find that the integral is $\lambda$ independent, and hence this term vanishes.

For the remaining (gauge-invariant) term in Eq. (3.52), we insert the expression

$$1 = \int_0^\infty \frac{d\lambda}{\lambda}\delta\left(1 - \frac{z_1+z_2}{\lambda}\right) \tag{3.57}$$

which greatly simplifies the formula to (after putting $z_i \to \lambda z_i$)

$$\bar{I}_{\mu\nu}(q) = \frac{2i\alpha}{\pi}\left(q_\mu q_\nu - g_{\mu\nu}q^2\right)\int_0^\infty dz_1\,dz_2\,z_1 z_2 \delta(1 - z_1 - z_2) \cdot \int_0^\infty \frac{d\lambda}{\lambda}\sum_i c_i \exp\left[i\lambda\left(q^2 z_1 z_2 - M_i^2 + i\epsilon\right)\right] \tag{3.58}$$

We need only one regulator mass and put $c_1 = -c_2 = +1$, $c_n = O$ $(n \geqslant 3)$, $M_2 = M$, and $M_n = O$ $(n \geqslant 3)$. Then

$$\bar{I}_{\mu\nu}(q) = I_{\mu\nu}\left(q, m^2\right) - I_{\mu\nu}\left(q, M^2\right) \tag{3.59}$$

$$= \frac{2i\alpha}{\pi}\left(q_\mu q_\nu - g_{\mu\nu}q^2\right)\int_0^1 dz\, z(1-z)$$

$$\cdot \int_0^\infty \frac{d\lambda}{\lambda} e^{i\lambda q^2 z_1 z_2}\left[e^{-\lambda(m^2-i\epsilon)} - e^{-i\lambda(M^2-i\epsilon)}\right] \tag{3.60}$$

Now use (consider $d/da$, $d/dA$, etc.)

$$\int_0^\infty \frac{d\lambda}{\lambda}\left[e^{-i\lambda(a-i\epsilon)} - e^{-i\lambda(A-i\epsilon)}\right] = \ln\frac{A}{a} \tag{3.61}$$

to arrive at

$$I_{\mu\nu}(q) = \frac{2i\alpha}{\pi}\left(q_\mu q_\lambda - q^2 g_{\mu\nu}\right)\int_0^1 dz\, z(1-z)\ln\frac{M^2 - q^2 z(1-z)}{m^2 - q^2 z(1-z)} \tag{3.62}$$

$$= \frac{2i\alpha}{\pi}\left(q_\mu q_\lambda - q^2 g_{\mu\nu}\right)\int_0^1 dz\, z(1-z)$$

$$\cdot\left[\ln\frac{M^2}{m^2} + \frac{q^2}{m^2}z(1-z) + \text{higher order in } q^2\right] \tag{3.63}$$

$$= \frac{i\alpha}{3\pi}\left(q_\mu q_\lambda - q^2 g_{\mu\nu}\right)\left(\ln\frac{M^2}{m^2} + \frac{q^2}{5m^2} + \cdots\right) \tag{3.64}$$

Thus gauge invariance ensures absence of the quadratic divergence, but the logarithmic divergence remains. Before interpreting the results, let us see how much easier it is to obtain Eq. (3.64), by dimensional regularization.

Starting again at Eq. (3.43), we use a Feynman parameter formula to obtain

$$I_{\mu\nu}(q) = -\frac{4e^2}{(2\pi)^4}\int_0^1 dx \int \frac{d^n k}{(k^2 + 2k\cdot Q - M^2)^2} T_{\mu\nu}(k,q) \tag{3.65}$$

where

$$Q_\mu = -x q_\mu \tag{3.66}$$

$$M^2 = m^2 - x q^2 \tag{3.67}$$

Using Eq. (3.45) for the Dirac trace, the dimensional regularization formulas give

$$\begin{aligned} I_{\mu\nu}(q) = &-\frac{4e^2}{(2\pi)^4}\int_0^1 \frac{dx}{(Q^2+M^2)^{2-n/2}} \\ &\cdot\left\{ g_{\mu\nu}(Q^2+M^2)\left(\frac{n}{2}-1\right)\Gamma\left(1-\frac{n}{2}\right) + \Gamma\left(2-\frac{n}{2}\right)\right. \\ &\left.\cdot\left[2Q_\mu Q_\nu + (q_\mu Q_\nu + q_\nu Q_\mu) + g_{\mu\nu}\left(-Q^2 - q\cdot Q + m^2\right)\right]\right\} \end{aligned} \tag{3.68}$$

Now because of the identity

$$\left(\frac{n}{2}-1\right)\Gamma\left(1-\frac{n}{2}\right)=-\Gamma\left(2-\frac{n}{2}\right) \tag{3.69}$$

we see immediately that the quadratic divergence (pole at $n = 2$) vanishes! Inserting Eqs. (3.66) and (3.67) into Eq. (3.68) now gives

$$I_{\mu\nu} = \frac{2i\alpha}{\pi}(q_\mu q_\nu - g_{\mu\nu}q^2)\Gamma\left(2-\frac{n}{2}\right)$$

$$\cdot\int dx \frac{x(1-x)}{[m^2-q^2x(1-x)]^{2-(n/2)}} \tag{3.70}$$

$$= \frac{i\alpha}{3\pi}(q_\mu q_\nu - g_{\mu\nu}q^2)\left[\frac{2}{4-n}+\frac{1}{5}\frac{q^2}{m^2}+O(q^4)\right] \tag{3.71}$$

That this agrees with Eq. (3.64) is evident when we identify the residue of the pole $(4-n)^{-1}$ with the coefficient of $\ln M$ in the Pauli–Villars cutoff method. The derivation is, however, obviously simpler.

For the reader unfamiliar with QED renormalization, we add some remarks about the result. The photon propagator is modified to

$$-\frac{ig_{\mu\nu}}{q^2+i\epsilon}\left[1-\frac{\alpha}{3\pi}\frac{2}{4-n}-\frac{\alpha}{15\pi}\frac{q^2}{m^2}+O(q^4,\alpha^2)\right] \tag{3.72}$$

The $q^2$-independent part is absorbed into the renormalization constant

$$Z_3 = 1-\frac{\alpha}{3\pi}\frac{2}{4-n}+O(\alpha^2) \tag{3.73}$$

The renormalized charge is given by

$$\alpha_R = z_3\alpha_B \tag{3.74}$$

(we are here anticipating the identity $Z_1 = Z_2$ derived below). The renormalized value $\alpha_R = (137.036)^{-1}$ is the measured fine-structure constant, while the bare value $\alpha_B$ is not observable. We also have the field renormalization for the photon

$$A_\mu = \sqrt{Z_3}A_{\mu R} \tag{3.75}$$

The $q^2$-dependent Uehling term is also very interesting. For low $q^2$ the propagator in momentum space is modified to

$$\frac{1}{q^2}\left[1-\frac{\alpha_R}{15\pi}\frac{q^2}{m^2}+O(\alpha_R^2,q^4)\right] \tag{3.76}$$

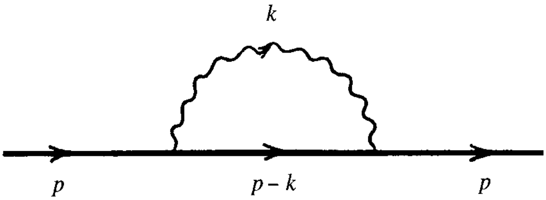



**Figure 3.2** Electron self-mass.

In coordinate space, this modifies the Coulomb potential to

$$\frac{e_R^2}{4\pi r} + \frac{4\alpha_R^2}{15m^2}\delta^3(\mathbf{x}) \tag{3.77}$$

The extra term affects the energy levels of the hydrogen atom, and in particular gives a substantial contribution to the Lamb shift [21]. The levels $2S_{1/2}$ and $2P_{1/2}$ are degenerate except for field quantization effects, which lift the $2S_{1/2}$ level by 1057.9 MHz (megahertz; $10^6$ cycles/second) relative to $2P_{1/2}$. The second term in Eq. (3.77) contributes an amount (−27.1 MHz) to this, and since the agreement [22] between theory and experiment is better than 0.1 MHz for the Lamb shift, the value for the vacuum polarization graph is well checked. The successful prediction of the Lamb shift is one of the triumphs of QED (although not the most accurate check, as will be seen in Example 3).

**Example 2.** Consider the electron self-mass diagram, Fig. 3.2. We shall again compare Pauli–Villars with dimensional regularization (for the last time; future example will be done only dimensionally). The electron propagator is replaced according to

$$\frac{i}{\not p - m} \to \frac{i}{\not p - m}\left[-\Sigma(p)\right]\frac{i}{\not p - m} \tag{3.78}$$

$$\Sigma(p) = \frac{-ie^2}{(2\pi)^4}\int d^4k \frac{1}{k^2 - \lambda^2 + i\epsilon}\gamma_\nu \frac{\not p - \not k + m}{(p-k)^2 - m^2 + i\epsilon}\gamma_\nu \tag{3.79}$$

where $\lambda$ is a (small) photon mass to accommodate infrared divergences.

Now use Eq. (3.46), the change of variable (3.48), and the integrals (3.51) to arrive at

$$\Sigma(p) = \frac{\alpha}{2\pi}\int_0^\infty \frac{dz_1\,dz_2}{(z_1+z_2)^2}\left(2m - \frac{\not p z_1}{z_1+z_2}\right) \cdot \exp\left[i\left(\frac{p^2 z_1 z_2}{z_1+z_2} - m^2 z_2 - \lambda^2 z_1\right)\right] \tag{3.80}$$

To make the integral converge, we introduce, a la Pauli–Villars, a heavy photon of mass $\Lambda$ and define

$$\bar{\Sigma}(p) = \Sigma(p, m, \lambda) - \Sigma(p, m, \Lambda) \tag{3.81}$$

Using the same scaling trick as in Example 1, we rescale $z_i \to \gamma z_i$ and insert

$$1 = \int_0^\infty \frac{d\gamma}{\gamma} \delta\left(1 - \frac{z_1 + z_2}{\gamma}\right) \tag{3.82}$$

to find

$$\begin{aligned} \bar{\Sigma}(p) &= \frac{\alpha}{2\pi} \int_0^1 dz \big[2m - \not{p}(1-z)\big] \\ &\quad \cdot \int_0^\infty \frac{d\gamma}{\gamma} \big[\exp\{i\gamma\big[p^2 z(1-z) - m^2 z - \lambda^2(1-z) + i\epsilon\big]\} \\ &\quad - (\lambda \leftrightarrow \Lambda)\big] \end{aligned} \tag{3.83}$$

Using Eq. (3.61) gives

$$\bar{\Sigma}(p) = \frac{\alpha}{2\pi} \int_0^1 dz \big[2m - \not{p}(1-z)\big] \ln \frac{\Lambda^2(1-z)}{\lambda^2(1-z) + m^2 z - p^2 z(1-z)} \tag{3.84}$$

Ensuring that the finite part vanish on mass shell $p^2 = m^2$ then requires us to write this as

$$\bar{\Sigma}(p) = \frac{\alpha}{2\pi} \int_0^1 dz \big[2m - \not{p}(1-z)\big] \left[\ln \frac{\Lambda^2}{m^2} + \ln \frac{m^2 z}{m^2 - p^2(1-z)}\right] \tag{3.85}$$

Here we have assumed that $\lambda^2 < (p^2 - m^2)$ and have subtracted $\ln[(1-z)/z^2]$ from the logarithm in Eq. (3.84) to fix the finite part uniquely.

We now need integrals (putting $p^2/m^2 = t$), namely,

$$\int dz \ln \frac{z}{1 - t + tz} = z \ln z - \frac{1 - t + tz}{t} \ln(1 - t + tz) \tag{3.86}$$

$$\begin{aligned} \int dz(1-z) \ln \frac{z}{1-t+tz} &= \left(z - \frac{1}{2} z^2\right) \ln \frac{z}{1-t+tz} \\ &\quad + \frac{1-t}{2t}\left[z - \frac{1+t}{t} \ln(1-t+tz)\right] \end{aligned} \tag{3.87}$$

Using these results in Eq. (3.85) gives

$$\begin{aligned} \bar{\Sigma}(p) &= \frac{3\alpha m}{4\pi} \ln \frac{\Lambda^2}{m^2} - \frac{\alpha}{2\pi}(\not{p} - m) \ln \frac{\Lambda^2}{m^2} \\ &\quad + \frac{\alpha m}{\pi} \frac{m^2 - p^2}{p^2} \ln \frac{m^2 - p^2}{m^2} \\ &\quad - \frac{\alpha \not{p}}{4\pi} \frac{m^2 - p^2}{p^2} \left(1 + \frac{m^2 + p^2}{p^2} \ln \frac{m^2 - p^2}{m^2}\right) \end{aligned} \tag{3.88}$$

Next to a free-particle spinor $\not{p} = m$ by the Dirac equation and with $p^2 \simeq m^2$,

$$\bar{\Sigma}(p) \to \frac{3\alpha m}{4\pi} \ln \frac{\Lambda^2}{m^2} - \frac{\alpha}{4\pi}(\not{p} - m)\left( \ln \frac{\Lambda^2}{m^2} + 4 \ln \frac{m^2 - p^2}{m^2} \right) \tag{3.89}$$

We absorb the divergences into the multiplicative renormalization constant

$$Z_2 = 1 - \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + O(\alpha^2) \tag{3.90}$$

and an additive mass renormalization

$$\delta m = \frac{3\alpha m}{4\pi} \ln \frac{\Lambda^2}{m^2} + O(\alpha^2) \tag{3.91}$$

To this order the propagator may now be written

$$\frac{i}{\not{p} - m} + \frac{i}{\not{p} - m}\left[-i\Sigma(p)\right]\frac{i}{\not{p} - m} = \frac{i}{\not{p} - m - \Sigma(p)} + O(\alpha^2) \tag{3.92}$$

$$= \frac{iZ_2}{\not{p} - m - \delta m} + O(\alpha^2) \tag{3.93}$$

The physical electron mass is $(m + \delta m)$, $m$ itself being unobservable. The electron field is renormalized according to

$$\psi = \sqrt{Z_2}\psi_R \tag{3.94}$$

As promised, we now present a simpler derivation of $\Sigma(p)$ using dimensional methods. From Eq. (3.79), using $\gamma_\nu \not{p} \gamma_\nu = -2\not{p}$ and Feynman parameters, we have

$$\Sigma(p) = \frac{-2ie^2}{(2\pi)^4} \int_0^1 dx \int d_k^n \frac{-\not{p} + \not{k} + 2m}{(k^2 + 2k \cdot Q - M^2)^2} \tag{3.95}$$

with

$$Q_\mu = -x p_\mu \tag{3.96}$$

$$M^2 = (m^2 - r^2)x \tag{3.97}$$

Dimensional regularization now gives

$$\Sigma(p) = \frac{-2ie^2}{(2\pi)^4} \frac{i\pi^2}{\Gamma(2)} \Gamma\left(2 - \frac{n}{2}\right) \int_0^1 \frac{dx(2m - \not{p} - \not{Q})}{(Q^2 + M^2)^{2-n/2}} \tag{3.98}$$

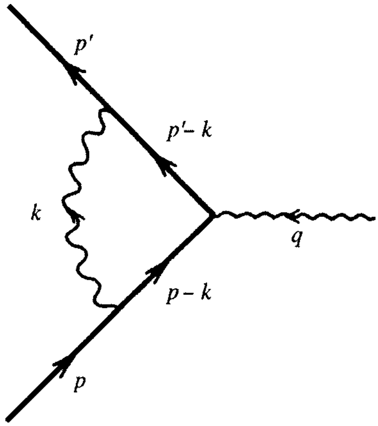



**Figure 3.3** QED vertex correction.

Expanding the denominator as

$$\left(Q^2 + M^2\right)^{-2+n/2} = 1 + \left(\frac{n}{2} - 2\right) \cdot \ln\left[m^2 x - p^2 x(1-x)\right] + O(n-4)^2 \tag{3.99}$$

then gives

$$\Sigma(p) = \frac{3\alpha m}{4\pi}\frac{2}{4-n} - \frac{\alpha}{4\pi}(\not{p} - m)\frac{2}{4-n} + \frac{\alpha}{2\pi}\int_0^1 dx\left[2m - \not{p}(1-x)\right]\ln\frac{m^2x^2}{m^2x - p^2x(1-x)} \tag{3.100}$$

where we have decreed, as before, that the nonpole (finite) part vanish at the renormalization point, $p^2 = m^2$. This formula agrees precisely with Eqs. (3.85) and (3.89).

**Example 3.** Consider the QED vertex correction of Fig. 3.3. The corresponding amplitude is

$$\Lambda_\mu(p', p) = (-ie)^2 \int \frac{d^4k}{(2\pi)^4}\left(\frac{-i}{k^2 - \lambda^2 + i\epsilon}\gamma_\nu \frac{i}{\not{p} - \not{k} - m + i\epsilon} \cdot \gamma_\mu \frac{i}{\not{p} - \not{k} - m + i\epsilon}\gamma_\nu\right) \tag{3.101}$$

This is understood to be sandwiched between spinors $\bar{u}(p') \cdots u(p)$, so that $\not{p}'$ on the left and $\not{p}$ on the right may be replaced by $m$ according to the Dirac equation.

Writing the Dirac factor as

$$t_\mu = \gamma_\nu(\not{p}' - \not{k} + m)\gamma_\mu(\not{p} - \not{k} + m)\gamma_\nu \tag{3.102}$$

we have, using Feynman parameters,

$$\Lambda_\mu(p', p) = \frac{-2ie^2}{(2\pi)^4} \int_0^1 dx \int_0^{1-x} d^4k \frac{t_\mu}{(k^2 + 2k \cdot Q - M^2)^3} \tag{3.103}$$

with

$$Q_\mu = -xp'_\mu - yp_\mu \tag{3.104}$$

$$M^2 = \lambda^2(1 - x - y) \tag{3.105}$$

$$Q^2 + M^2 = m^2(x + y)^2 + \lambda^2(1 - x - y) - q^2xy \tag{3.106}$$

Applying dimensional regularization corresponds to the recipe

$$k_\alpha k_\beta \to Q_\alpha Q_\beta \quad \text{(aside from divergent piece)} \tag{3.107}$$

$$k_\alpha \to -Q_\alpha \tag{3.108}$$

$$1 \to 1 \tag{3.109}$$

Then it is straightforward but tedious to find that (putting $\not{p}' = m$ on left, $\not{p} = m$ on right)

$$E_\mu = \gamma_\nu(\not{p}' + \not{Q} + m)\gamma_\mu(\not{p} + \not{Q} + m)\gamma_\nu \tag{3.110}$$

$$= A\gamma_\mu m^2 + B\gamma_\mu q^2 + Cmp_\mu + DMp'_\mu \tag{3.111}$$

with

$$A = 4 - 8(x + y) + 2(x + y)^2 \tag{3.112}$$

$$B = -2(1 - x)(1 - y) \tag{3.113}$$

$$C = -4xy + 4x - 4y^2 \tag{3.114}$$

$$D = -4xy + 4y - 4x^2 \tag{3.115}$$

Thus $\Lambda_\mu(p', p)$ contains the purely finite part (here we may put $n = 4$ immediately),

$$\Lambda_\mu^{(1)}(p', p) = \frac{-2ie^2}{(2\pi)^4} \frac{-i\pi^2}{\Gamma(3)} \int \frac{dx\, dy E_\mu}{Q^2 + M^2} \tag{3.116}$$

There is also the log divergent part of $k_\alpha k_\beta$ giving

$$\Lambda_\mu = \Lambda_\mu^{(1)} + \Lambda_\mu^{(2)} \tag{3.117}$$

with

$$\Lambda_\mu^{(2)}(p', p) = \frac{-2ie^2}{(2\pi)^4} \frac{-i\pi^2}{\Gamma(3)} \int \frac{dx\,dy}{(Q^2 + M^2)^{2-n/2}} \left(-\frac{1}{2}\right) \Gamma\left(2 \frac{n}{2}\right) (\gamma_\nu \gamma_\alpha \gamma_\mu \gamma_\alpha \gamma_\nu) \tag{3.118}$$

Let us look at the divergent term only first, since it is relevant to the Ward identity. Using

$$\gamma_\nu \gamma_\alpha \gamma_\mu \gamma_\alpha \gamma_\nu = 4\gamma_\mu \tag{3.119}$$

we have ($\alpha = e^2/4\pi$)

$$\Lambda_\mu^{(2)}(p', p) \sim \frac{\alpha}{4\pi} \gamma_\mu \frac{2}{4-n} \tag{3.120}$$

We define a renormalization constant $Z_1$ by

$$\bar{u}(p) \Lambda_\mu(p, p) u(p) = \left( \frac{1}{Z_1} - 1 \right) \bar{u}(p) \gamma_\mu u(p) \tag{3.121}$$

Thus

$$Z_1 = 1 - \frac{\alpha}{4\pi} \frac{2}{4-n} + O\left(\alpha^2\right) \tag{3.122}$$

The reader will notice that this coincides precisely with the expression for $Z_2$ given in Eq. (3.90). This is no accident and is the Ward–Takahashi identity [23–25] for QED. It holds to all orders in perturbation theory, namely,

$$Z_1 = Z_2 \tag{3.123}$$

At order $\alpha$ it is easily derived by noticing that

$$\frac{\partial}{\partial p_\mu} \frac{1}{\not{p} - \not{k} - m + i\epsilon} = \frac{-1}{\not{p} - \not{k} - m + i\epsilon} \gamma_\mu \frac{1}{\not{p} - \not{k} - m + i\epsilon} \tag{3.124}$$

Thus comparing Eq. (3.79) for $\Sigma(p)$ with Eq. (3.101) for $\Lambda(p', p)$, we see immediately that

$$\frac{d}{dp_\mu} \Sigma(p) = -\Lambda_\mu(p, p) \tag{3.125}$$

Since we showed in Example 2 that the divergent part of $\Sigma(p)$ is given by

$$\Sigma(p) \sim \frac{\alpha}{4\pi} (\not{p} - m) \frac{2}{4-n} + \frac{3\alpha m}{4\pi} \frac{2}{4-n} \tag{3.126}$$

$$= (\not{p} - m) \left( \frac{1}{Z_2} - 1 \right) + \delta m \tag{3.127}$$

we see that

$$\lambda_\mu(p, p) = -\frac{\partial}{\partial p_\mu}\Sigma(p) \tag{3.128}$$

$$\approx \gamma_\mu\left(\frac{1}{Z_2} - 1\right) \tag{3.129}$$

$$= \gamma_\mu\left(\frac{1}{Z_1} - 1\right) \tag{3.130}$$

Hence the identity (3.123) follows.

Thus there are only three independent renormalization constants in QED: $Z_3$ for the photon field, $Z_2$ for the electron field, and $\delta m$ for the electron mass. The multiplicative renormalization of the charge depends only on $Z_3$, as indicated without proof in Example 2. We can now verify this at order $\alpha$ by considering all the diagrams that contribute. They are depicted in Fig. 3.4; these diagrams are self-explanatory except Fig. 3.4*e* and *f*, which are depictions of the counterterms $(-\delta m)$ associated with Fig. 3.4*c* and *d*, respectively. In the limit $p'_\mu = p_\mu$, the seven contributions are listed below [there is a common factor $(-ie)$].

| | |
|---|---|
| (a) | $\gamma_\mu$ |
| (b) | $\gamma_\mu(Z_3 - 1)$ |
| (c) + (e) | $-\left(\frac{1}{Z_2} - 1\right)\gamma_\mu$ |
| (d) + (f) | $-\left(\frac{1}{Z_2} - 1\right)\gamma_\mu$ |
| (g) | $\gamma_\mu\left(\frac{1}{Z_2} - 1\right)$ |

The external lines require field renormalization factors $[(\sqrt{Z_2})^2\sqrt{Z_3}]^{-1}$.

The resultant sum is thus given by

$$\frac{\gamma_\mu}{Z_2\sqrt{Z_3}}\left[1 + (Z_3 - 1) - 2\left(\frac{1}{Z_2} - 1\right) + \left(\frac{1}{Z_1} - 1\right)\right]$$

$$= \frac{\gamma_\mu}{Z_2\sqrt{Z_3}}\left(1 + Z_3 - \frac{1}{Z_2}\right) \tag{3.131}$$

$$= \frac{\gamma_\mu}{Z_2\sqrt{Z_3}}\frac{1 + (Z_3 - 1)}{1 + [(1/Z_2) - 1]} + O(\alpha^2) \tag{3.132}$$

$$= \sqrt{Z_3}\gamma_\mu + O(\alpha^2) \tag{3.133}$$

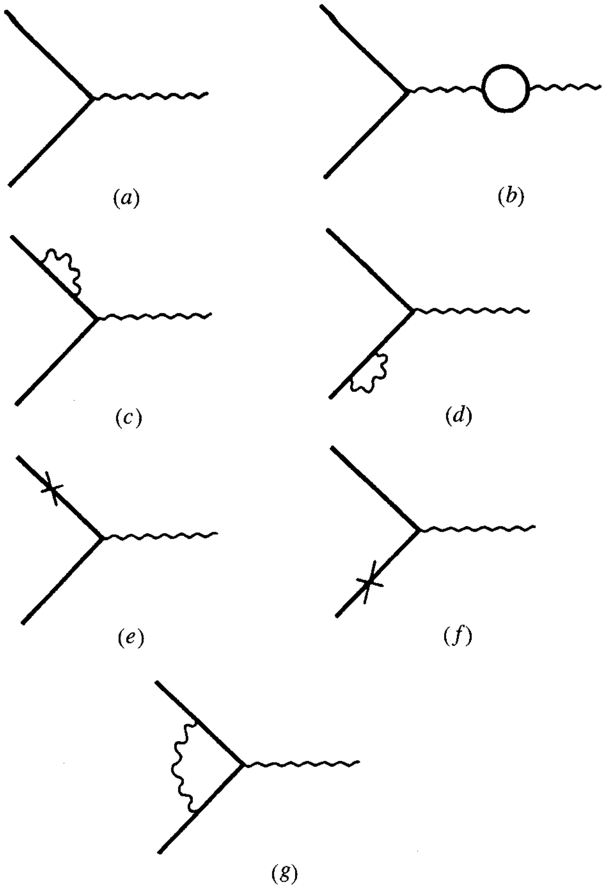


**Figure 3.4** Order $\alpha$ contributions to charge renormalization.

This shows that the renormalization constant $Z_1 = Z_2$ cancels out and justifies Eq. (3.74).

So far we have studied only the divergent part of $\Lambda_\mu(p', p)$. There is a finite part proportional to $\gamma_\mu$ coming from $\Lambda_\mu^{(2)}$ in Eq. (3.118) and from the terms $A$ and $B$ of Eq. (3.111) substituted into $\Lambda_\mu^{(1)}$ of Eq. (3.116); this term will not be worked through, since it is not of special interest.

Of very great interest, for checking QED, are the remaining finite terms, which are provided by $C$ and $D$ in Eq. (3.111) substituted into $\Lambda_\mu^{(1)}$ of Eq. (3.116). The dimensional regularization again simplifies the calculation. We have a contribution to $\Lambda_\mu(p', p)$ of

$$-\frac{2ie^2}{(2\pi)^4}\frac{-i\pi^2}{\Gamma(3)}\int\frac{dx\,dy}{m^2(x+y)^2}m\big[-\big(4xy+4x-4y^2\big)p_\mu$$
$$+\big(-4xy+4y-4x^2\big)p'_\mu\big]$$

$$= -\frac{e^2}{4\pi^2 m}\int \frac{dx\,dy}{(x+y)^2}\big[\big(-xy+x-y^2\big)p_\mu + \big(-xy+y-x^2\big)p'_\mu\big]$$

$$+\,O\big(q^2\big) \tag{3.134}$$

$$= -\frac{\alpha}{4\pi m}(p+p')_\mu + O\big(q^2\big) \tag{3.135}$$

where we used the integrals

$$\int_0^1 dx \int_0^{1-x} dy \frac{1}{(x+y)^2}\{x, y, x^2, y^2, xy\} = \left\{\frac{1}{2}, \frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{12}\right\} \tag{3.136}$$

Sandwiched between spinors, we have the identity

$$\bar{u}(p')\frac{(p+p')_\mu}{2m}u(p) \equiv \bar{u}(p')\left(\gamma_\mu - \frac{i\sigma_{\mu\nu}q_\nu}{2m}\right)u(p) \tag{3.137}$$

where

$$\sigma_{\mu\nu} = \frac{1}{2}[\gamma_\nu, \gamma_\nu] \tag{3.138}$$

Thus the contribution (3.135) to $\Lambda_\mu(p', p)$ may be rewritten

$$-\frac{\alpha}{2\pi}\gamma_\mu + \frac{i\alpha}{4\pi m}\sigma_{\mu\nu}q_\nu \tag{3.139}$$

The first term cancels against the other contributions proportional to $\gamma_\mu$ and of order $\alpha$ (recall that only the $q^2 \to 0$ limit is needed). Combined with the Born term, we therefore have the static limit

$$\bar{u}(p')\left(\gamma_\mu + \frac{i\alpha}{4\pi m}\sigma_{\mu\nu}q_\nu\right)u(p)$$
$$= \bar{u}(p')\left[\frac{(p+p')_\mu}{2m} + \left(1+\frac{\alpha}{2\pi}\right)\frac{i\sigma_{\mu\nu}q_\nu}{2m}\right]u(p) \tag{3.140}$$

The $\alpha/2\pi$ coefficient provides the lowest-order correction to the magnetic moment of the electron whose gyromagnetic ratio $g$ is thus given by [19, Chap. 1]

$$\frac{1}{2}g = 1 + \frac{\alpha}{2\pi} + O\big(\alpha^2\big) \tag{3.141}$$

This correction was first calculated by Schwinger [26] in 1948 and found to agree with the experimental value of the anomalous moment determined in 1947 by Kusch and Foley [27, 28]. Since then the theoretical value has been computed to order $\alpha^4$ giving (the numbers quoted below are from Refs. [29, 30])

$$a_e^{\rm th} = \frac{1}{2}(g_e - 2) \tag{3.142}$$

$$= 0.5\left(\frac{\alpha}{\pi}\right) - 0.32848\left(\frac{\alpha}{\pi}\right)^2 + 1.49\left(\frac{\alpha}{\pi}\right)^3 \tag{3.143}$$

$$= 0.001\,159\,652\,188\,3(7) \tag{3.144}$$

The experimental value is

$$a_e^{\rm expt} = 0.001\,159\,657 \tag{3.145}$$

agreeing within errors to about 1 in $10^{10}$ for $a = \frac{1}{2}(g - 2)$ and to 1 in $10^{13}$ for $g$ itself.

**Example 4.** As a final example of dimensional regularization technique,[1] and as a bridge to Section 3.3, we calculate the process $S \to \gamma\gamma$, where $S$ is a scalar and both photons may be off-mass shell. The interaction is taken as

$$L = ig\bar{\psi}\psi\phi - e\bar{\psi}\gamma_\mu\psi A_\mu \tag{3.146}$$

and the two diagrams are depicted in Fig. 3.5.

Taking the amplitude for Fig. 3.5*a* as $T_{\mu\nu}(p_1, p_2)$, the sum of the two diagrams will be

$$I_{\mu\nu}(p_1, p_2, m) = T_{\mu\nu}(p_1, p_2) + T_{\mu\nu}(p_2, p_1) \tag{3.147}$$

It will turn out that $T_{\mu\nu}$ alone satisfies Bose symmetry, so that adding the crossed diagram, as in Eq. (3.147), merely multiplies $T_{\mu\nu}$ by a factor of 2 (this is an accident for the present special case and should not be expected in general).

The Feynman rules give

$$\begin{aligned} T_{\mu\nu} &= -(i)^3 g(-ie)^2 \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2 + i\epsilon} \\ &\quad \cdot \frac{1}{(k - p_1)^2 - m^2 + i\epsilon} \frac{1}{(k + p_2)^2 - m^2 + i\epsilon} 4t_{\mu\nu} \end{aligned} \tag{3.148}$$

with the Dirac trace $t_{\mu\nu}$ given by

$$4t_{\mu\nu} = \mathrm{Tr}\left[(\not{k} + m)\gamma_\mu(\not{k} - \not{p}_1 + m)(\not{k} + \not{p}_2 + m)\gamma_\nu\right] \tag{3.149}$$

This is easily evaluated as

$$\begin{aligned} t_{\mu\nu} &= m\left[4k_\mu k_\nu + 2(k_\mu p_{2\nu} - p_{1\mu}k_\nu)\right. \\ &\quad \left. - (p_{1\mu}p_{2\nu} - p_{2\mu}p_{1\nu}) + g_{\mu\nu}\left(m^2 - k^2 - p_1 \cdot p_2\right)\right] \end{aligned} \tag{3.150}$$

1) Here the technique is applied, with great convenience, to a finite amplitude.

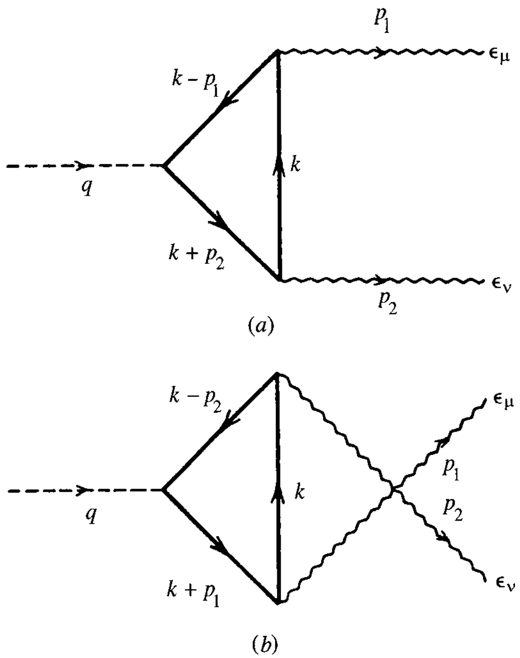



**Figure 3.5** Diagrams for $S \to \gamma\gamma$.

Introduce Feynman parameters to obtain

$$T_{\mu\nu} = \frac{8ige^2}{(2\pi)^4} \int_0^1 dx \int_0^{1-x} dy \int \frac{d^n k}{(k^2 + 2k \cdot Q - M^2)^3} t_{\mu\nu} \tag{3.151}$$

where

$$Q_\mu = -xp_{1\mu} + yp_{2\mu} \tag{3.152}$$

$$M^2 = -xp_1^2 - yp_2^2 + m^2 \tag{3.153}$$

Dimensional regularization now gives

$$\begin{aligned} T_{\mu\nu} = {} & \frac{8ige^2}{(2\pi)^4} \frac{i\pi^2}{\Gamma(3)} \int_0^1 dx \int_0^{1-x} dy \int \frac{1}{(-Q^2 - M^2)^{3-n/2}} \\ & \cdot \Big\{ \big[ 4Q_\mu Q_\nu - 2Q_\mu p_{2\nu} + 2p_{1\mu} Q_\nu - p_{1\mu} p_{2\nu} + p_{2\mu} p_{1\nu} \\ & + g_{\mu\nu} (m^2 - p_1 \cdot p_2 - Q^2) \big] \Gamma\left(3 - \frac{n}{2}\right) \\ & + g_{\mu\nu} (-Q^2 - M^2) \left(2 - \frac{1}{2}n\right) \Gamma\left(2 - \frac{1}{2}n\right) \Big\} \end{aligned} \tag{3.154}$$

Since, in the final term of the braces, one has

$$\left(2 - \frac{1}{2}n\right)\Gamma\left(2 - \frac{1}{2}n\right) = \Gamma\left(3 - \frac{1}{2}n\right) \tag{3.155}$$

we see that the amplitude is finite. Substituting Eqs. (3.152) and (3.153) into Eq. (3.154) and using Eq. (3.155) now gives

$$\begin{aligned} T_{\mu\nu} = {} & \frac{ge^2 m}{4\pi^2}\int_0^1 dx \\ & \cdot \int_0^{1-x} dy \frac{1}{p_1^2 x(1-x) + p_2^2 y(1-y) + 2p_1 \cdot p_2 xy - m^2} \\ & \cdot \big[p_{1\mu}p_{1\nu}\big(4x^2 - 2x\big) + p_{2\mu}p_{2\nu}\big(4y^2 - 2y\big) \\ & + p_{1\mu}p_{2\nu}(-4xy + 2x + 2y - 1) \\ & + p_{2\mu}p_{1\nu}(1 - 4xy) + g_{\mu\nu}\big(p_1^2 x + p_2^2 y - 2p_1^2 x^2 - 2p_1^2 y^2 \\ & + 4p_1 \cdot p_2 xy - p_1^2 p_2\big)\big] \end{aligned} \tag{3.156}$$

This is seen to satisfy

$$T_{\mu\nu}(p_1, p_2) = T_{\nu\mu}(p_2, p_1) \tag{3.157}$$

as anticipated above.

Let us adopt the shorthand (for the rest of this section *only*)

$$\begin{aligned} \iint \equiv {} & \frac{ge^2 m}{2\pi^2}\int_0^1 dx \\ & \cdot \int_0^{1-x} dy \frac{1}{p_1^2 x(1-x) + p_2^2 y(1-y) + 2p_1 \cdot p_2 xy - m^2} \end{aligned} \tag{3.158}$$

Then we define the amplitudes

$$A = \iint (1 - 4xy) \tag{3.159}$$

$$B_1 = \frac{1}{p_2^2}\iint 2x(1 - 2x) \tag{3.160}$$

$$B_2 = \frac{1}{p_1^2}\iint 2y(1 - 2y) \tag{3.161}$$

$$B_3 = -\frac{1}{p_1 \cdot p_2}\iint (1 - 2x)(1 - 2y) \tag{3.162}$$

Then from Eq. (3.156) we find that

$$I_{\mu\nu} = A(p_{2\mu}p_{1\nu} - p_1 \cdot p_2 g_{\mu\nu}) + \frac{1}{2}(B_1 + B_2)p_1^2 p_2^2 g_{\mu\nu} - B_1 p_2^2 p_{1\mu} p_{1\nu} - B_2 p_1^2 p_{1\mu} p_{1\nu} + B_3 p_1 \cdot p_2 p_{1\mu} p_{2\nu} \tag{3.163}$$

We can now show that $B_1 = B_2 = B_3$ ($= B$, say), as follows. Consider

$$p_2^2 p_1 \cdot p_2 (B_1 - B_3) \propto \int_0^1 dx \int_0^{1-x} dy \frac{(1-2x)[2xp_1 \cdot p_2 + (1-2y)p_2^2]}{p_1^2 x(1-x) + p_2^2 y(1-y) + 2p_1 \cdot p_2 xy - m^2} \tag{3.164}$$

$$= -\int_0^\infty dz \int_0^1 dx(1-2x) \int_0^{1-x} dy \cdot \frac{\partial}{\partial y}\{\exp[-z(p_1^2 x(1-x) + p_2^2 y(1-y) + 2p_1 \cdot p_2 xy - m^2)]\} \tag{3.165}$$

Introducing $x' = x - \frac{1}{2}$, this expression is

$$p_2^2 p_1 \cdot p_2 (B_1 - B_3) \sim -\int_0^\infty dz\, e^{-M^2 z} \int_{-1/2}^{+1/2} dx'(-2x') \cdot \left\{\exp\left[-z(p_1+p_2)^2\left(\frac{1}{4} - x'^2\right)\right] - \exp\left[-zp_1^2\left(\frac{1}{4} - x'^2\right)\right]\right\} \tag{3.166}$$

$$= 0 \tag{3.167}$$

The integral vanishes since it is odd $x'$. Similarly, we may show that $B_2 - B_3$.

Thus, we have finally, from Eq. (3.163),

$$I_{\mu\nu} = A(p_{2\mu}p_{1\nu} - p_1 \cdot p_2 g_{\mu\nu}) + B\left(p_{1\mu}p_{2\nu}p_1 \cdot p_2 - p_1^2 p_{2\mu}p_{2\nu} - p_2^2 p_{1\mu}p_{1\nu} + p_1^2 p_2^2 g_{\nu\mu}\right) \tag{3.168}$$

This satisfies gauge invariance since

$$p_{1\mu} I_{\mu\nu} = p_{2\nu} I_{\mu\nu} = 0 \tag{3.169}$$

Of course, this property was guaranteed from the beginning, since the integral was convergent, but the explicit evaluation would be considerably more difficult using Pauli–Villars regularization.

## 3.3 Triangle Anomalies

We now turn to a curiously profound property of gauge theories with chiral fermions, which necessarily involve $\gamma_5$ in their gauge couplings. This is the case for example, the standard Glashow–Salam–Weinberg electroweak theory.

Recall the Dirac algebra

$$\{\gamma_\mu, \gamma_\nu\}_+ = 2g_{\mu\nu} \tag{3.170}$$

and the definition

$$\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3 \tag{3.171}$$

$$= \frac{1}{24}\epsilon_{\alpha\beta\gamma\delta}\gamma_\alpha\gamma_\beta\gamma_\gamma\gamma_\delta \tag{3.172}$$

so that

$$\{\gamma_5, \gamma_\mu\}_+ = 0 \tag{3.173}$$

One representation is the $4 \times 4$ matrices

$$\gamma_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.174}$$

$$\boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix} \tag{3.175}$$

$$\gamma_5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.176}$$

Dimensional regularization is not straightforwardly applicable to expressions involving $\gamma_5$ because the definition (3.172), in particular the tensor $\epsilon_{\alpha\beta\gamma\delta}$, is defined only in four dimensions (see, however, Refs. [31–36]).

As already emphasized, the renormalization procedure must be such that the renormalized quantities respect local gauge invariance. For Green's functions, this requirement is expressed by the Ward identities. In the Abelian case of quantum electrodynamics, these are the Ward–Takahashi identities [23–25] discussed above; in a non-Abelian theory, gauge invariance for Green's functions is most succinctly expressed by the Becchi–Rouet–Stora–Tyutin (BRST) identities [37], which summarize the earlier forms found by Taylor [38] and Slavnov [39]. The BRST transfor-

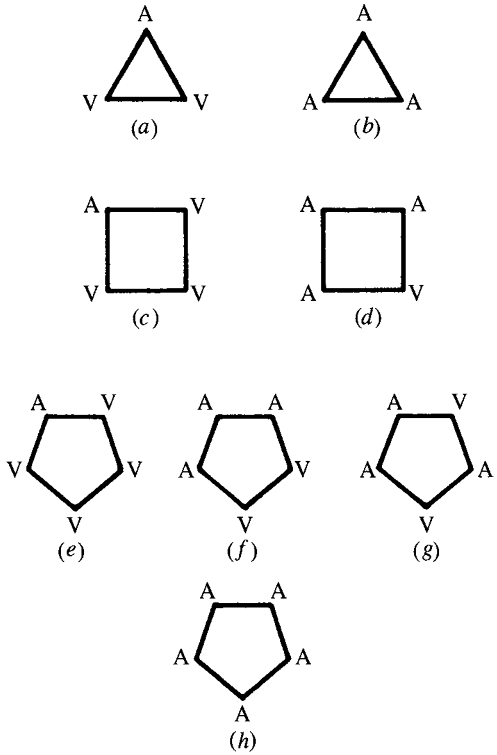



**Figure 3.6** Triangle, square, and pentagon diagrams for anomaly considerations. V, vector coupling; A, axial-vector coupling.

mation is discussed later; for present purposes we merely state the relevant Ward identities for the amplitudes considered.

In the presence of $\gamma_5$ couplings to the gauge field, there are axial vector Ward identities as well as vector identities. The problem of anomalies is that one finds triangular Feynman diagrams is one-loop order of perturbation theory that violate the axial Ward identities. Such anomalies must be canceled in order that the theory has a chance to be renormalizable.

The triangle anomaly has a long history, going back to work, by Steinberger [40] and Schwinger [41]. Considerably later, the anomaly was emphasized by others [42–45].

After proof of renormalizability of Yang–Mills theory in 1971, the relevance of anomaly cancellation to renormalizability was discussed by several authors [46–48] (on anomalies, see also Refs. [11], and [49–57]; for reviews, see Refs. [58] and [59]).

The Feynman graph giving rise to an anomaly is a triangular fermion loop with three gauge fields and with overall abnormal parity (i.e., one or three $\gamma_5$ couplings) as in Fig. 3.6*a* and *b*. The square and pentagon configuration indicated in Fig. 3.6*c* through *h* must be considered also [49]. One of two useful theorems [49] (see also Ref. [57]) is that once the AW triangle anomaly is canceled, so are all the others. The second useful theorem, due to Adler and Bardeen [52] (see also Ref. [58]) is

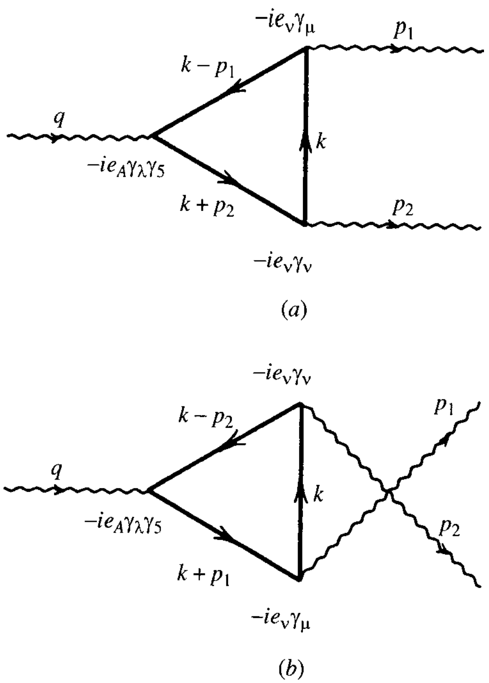



**Figure 3.7** (a) Triangle graph; (b) crossed graph.

that radiative corrections do not renormalize the anomaly. Thus we need consider only the simplest AVV graph.

The interaction Lagrangian will be taken as

$$L = -e_A \bar{\psi} \gamma_\mu \gamma_5 \psi Z^A_\mu - e_V \bar{\psi} \gamma_\mu \psi A_\mu \tag{3.177}$$

and, at first, we allow only one fermion species to propagate in the graph of Fig. 3.7*a*. The full amplitude requires adding the crossed graph of Fig. 3.7*b* according to

$$I_{\mu\nu\lambda}(p_1, p_2, m) = T_{\nu\mu\lambda}(p_1, p_2, m) + T_{\mu\nu\lambda}(p_2, p_1, m) \tag{3.178}$$

Feynman rules now give

$$\begin{aligned} T_{\mu\nu\lambda} &= (i)^3(-ie_A)(-ie_V)^2 \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2 + i\epsilon} \\ &\quad \cdot \frac{1}{(k - p_1)^2 + m^2 + i\epsilon} \frac{1}{(k + p_2)^2 - m^2 + i\epsilon} 4t_{\mu\nu\lambda} \end{aligned} \tag{3.179}$$

where the Dirac trace is

$$4t_{\mu\nu\lambda} = \mathrm{Tr}\left[(\not{k} + m)\gamma_\mu(\not{k} - \not{p}_1 + m)\gamma_\lambda\gamma_5(\not{k} + \not{p}_2 + m)\gamma_\nu\right] \tag{3.180}$$

The crucial point is that the integral in Eq. (3.179) is superficially *linearly* divergent, and therefore shifting the integration variable alters $T_{\mu\nu\lambda}$ by a finite amount. Such a shift is necessary to establish the Ward identities expected, namely

$$(p_1 + p_2)_\lambda I_{\mu\nu\lambda} = 0 \tag{3.181}$$

$$P_{1\mu} I_{\mu\nu\lambda} = 0 \tag{3.182}$$

$$P_{2\nu} I_{\mu\nu\lambda} = 0 \tag{3.183}$$

We will now show that if we impose the vector identities (3.182) and (3.183) (as is essential in quantum electrodynamics for charge conservation), then we do not maintain, in general, the axial-vector Ward identity, Eq. (3.181), unless extra conditions are met.

The linear divergence is independent of the fermion mass $m$, which we therefore set to zero, giving

$$T_{\mu\nu\lambda} = -e_A e_V^2 \int \frac{d^4k}{(2\pi)^4} \frac{\mathrm{Tr}[\not{k}\gamma_\mu(\not{k} - \not{p}_1)\gamma_\lambda\gamma_5(\not{k} + \not{p}_2)\gamma_\nu]}{k^2(k-p_1)^2(k+p_2)^2} \tag{3.184}$$

Computation of $T_{\mu\nu\lambda}$ reveals that it is Bose symmetric under the interchange $\{p_1, \mu\} \leftrightarrow \{p_2, \nu\}$, so that addition of the crossed term in Eq. (3.178) merely gives a factor 2. Thus identities (3.181) through (3.183) may be applied directly to $T_{\mu\nu\lambda}$.

Consider first

$$(p_1 + p_2)_\lambda T_{\mu\nu\lambda} = -e_A e_V^2 \int \frac{d^4k}{(2\pi)^4} \frac{\mathrm{Tr}[\not{k}\gamma_\mu(\not{k} - \not{p}_1)(\not{p}_1 + \not{p}_2)\gamma_5(\not{k} + \not{p}_2)\gamma_\nu]}{k^2(k-p_1)^2(k+p_2)^2} \tag{3.185}$$

Now rewrite

$$(\not{p}_1 + \not{p}_2)\gamma_5 \equiv -(\not{k} - \not{p}_1)\gamma_5 - \gamma_5(\not{k} + \not{p}_2) \tag{3.186}$$

to find that

$$(p_1 + p_2)_\gamma T_{\mu\nu\lambda} = \frac{e_A e_V^2}{(2\pi)^4}\left\{\int d^4k \frac{\mathrm{Tr}[\not{k}\gamma_\mu\gamma_5(\not{k} + \not{p}_2)\gamma_\nu]}{k^2(k+p_2)^2} + \int d^4k \frac{\mathrm{Tr}[\not{k}\gamma_\mu(\not{k} - \not{p}_1)\gamma_5\gamma_\nu]}{k^2(k-p_1)^2}\right\} \tag{3.187}$$

Now both terms in the braces are second-rank pseudotensors depending on only one four-momentum. No such tensor exists, so we would conclude that Eq. (3.181) is satisfied. (However, the reader must not stop here but read on!) Now consider

$$p_{1\nu} T_{\mu\nu\lambda} = \frac{-e_A e_V^2}{(2\pi)^4} \int d^4k \frac{\mathrm{Tr}[\not{k} - \not{p}_1(\not{k} - \not{p}_1)\gamma_\lambda\gamma_5(\not{k} + \not{p}_2)\gamma_\nu]}{k^2(k-p_1)^2(k+p_2)^2} \tag{3.188}$$

Let us change variable to $k'_\mu = (k + p_2)_\mu$ and be deliberately careless (temporarily), to obtain

$$p_{1\nu} T_{\mu\nu\lambda} = \frac{-e_A e_V^2}{(2\pi)^4} \int d^4 k \frac{\mathrm{Tr}[(\not{k} - \not{p}_2)\not{p}_1(\not{k}' - \not{p}_1 - \not{p}_2)\gamma_\lambda \gamma_5 \not{k}' \gamma_\nu]}{(k' - p_2)^2 (k' - p_1 - p_2)^2 k'^2} \tag{3.189}$$

Now rewrite

$$\not{p}_1 = -(\not{k}' - \not{p}_1 - \not{p}_2) + (\not{k} - \not{p}_2) \tag{3.190}$$

to obtain, apparently,

$$\begin{aligned} P_{1\mu} T_{\mu\nu\lambda} = & \frac{e_A e_V^2}{(2\pi)^4} \left\{ \int d^4 k' \frac{\mathrm{Tr}[(\not{k}' - \not{p}_2)\gamma_\lambda \gamma_5 k' \gamma_\nu]}{(k' - p_2)^2 k'^2} \right. \\ & \left. - \int d^4 k' \frac{\mathrm{Tr}[(\not{k}' - \not{p}_1 - \not{p}_2)\gamma_\lambda \gamma_5 \not{k}' \gamma_\nu]}{(k' - p_1 - p_2)^2 k'^2} \right\} \end{aligned} \tag{3.191}$$

which vanishes exactly as before.

By similar steps, putting $k'' = (k - p_1)$ and rewriting

$$\not{p}_2 = (\not{k}'' + \not{p}_1 + \not{p}_2) - (\not{k}'' + \not{p}_1) \tag{3.192}$$

we can apparently obtain

$$\begin{aligned} p_{2\nu} T_{\mu\nu\lambda} = & \frac{-e_A e_V^2}{(2\pi)^4} \left\{ \int d^4 k'' \frac{\mathrm{Tr}[(\not{k}'' + \not{p}_1)\gamma_\mu k'' \gamma_\lambda \gamma_5]}{(k'' + p_1)^2 k''^2} \right. \\ & \left. - \int d^4 k'' \frac{\mathrm{Tr}[\gamma_\mu \not{k}'' \gamma_\lambda \gamma_5 (\not{k}'' + \not{p}_1 + \not{p}_2)]}{k''^2 (k'' + p_1 + p_2)^2} \right\} \end{aligned} \tag{3.193}$$

Both terms in brackets are second-rank pseudotensors depending only on one four-vector, and hence vanish.

The fallacy in the argument above [i.e., in Eqs. (3.185) through (3.193)] is that shifting the integration variable in a linearly divergent integral changes the value of that integral by a finite amount. Define $S_{\mu\nu\lambda}$ by

$$T_{\mu\nu\lambda} = -e_A e_V^2 S_{\mu\nu\lambda} \tag{3.194}$$

so the linearly divergent piece is

$$S_{\mu\nu\lambda} = \frac{1}{(2\pi)^4} \int d^4 k \frac{\mathrm{Tr}(\not{k}\gamma_\mu \not{k} \gamma_\lambda \gamma_5 \not{k} \gamma_\nu)}{k^6} \tag{3.195}$$

If we shift to $k' = (k + a)$, by how much does $S_{\mu\nu\lambda}$ change? Consider the manipulations

$$\int d^4k f(k) = \int d^4k' f(k' - a) \tag{3.196}$$

$$= \int d^4k' f(k') - a_\rho \int d^4k \frac{\partial}{\partial k_\rho} f(k) + \cdots \tag{3.197}$$

If the original integral is linearly divergent, the second term in Eq. (3.197) is finite, since when converted to a surface term by Gauss's theorem, the integrand $f(k) \sim |k|^{-3}$ and the surface area $\sim |k|^3$.

In the present case, suppose that we shift the integration variable in Eq. (3.195) to $k'_\mu = (k+a)_\mu$; then we have

$$S'_{\mu\nu\lambda} = S_{\mu\nu\lambda} + C_{\mu\nu\lambda\rho} a_\rho \tag{3.198}$$

where

$$C_{\mu\nu\lambda\rho} = -\frac{1}{(2\pi)^4} \int d^4k \frac{\partial}{\partial k_\rho} \frac{\mathrm{Tr}(\not k \gamma_\mu \not k \gamma_\lambda \gamma_5 \not k \gamma_\nu)}{k^6} \tag{3.199}$$

$$= -\frac{1}{(2\pi)^4} \int d^4k \frac{\partial}{\partial k_\rho} \frac{\mathrm{Tr}(\gamma_5 \not k \gamma_\nu \not k \gamma_\mu \not k \gamma_\lambda)}{k^6} \tag{3.200}$$

Recall now the trace formula

$$\mathrm{Tr}(\gamma_5 \gamma_\alpha \gamma_\beta \gamma_\delta \gamma_\epsilon \gamma_\phi) = -4i \epsilon_{\delta\epsilon\phi k}(g_{k\alpha} g_{\beta\gamma} - g_{k\beta} g_{\alpha\gamma} + g_{k\gamma} g_{\alpha\beta})$$
$$+ 4i \epsilon_{\alpha\beta\gamma k}(g_{k\delta} g_{\epsilon\phi} - g_{k\epsilon} g_{\delta\phi} + g_{k\phi} g_{\delta\epsilon}) \tag{3.201}$$

Using this one finds from Eq. (3.200) that

$$C_{\mu\nu\lambda\rho} = -\frac{4i}{(2\pi)^2} \int d^4k \frac{\partial}{\partial k_\rho} \frac{\epsilon_{\mu\nu\lambda\epsilon} k_\epsilon}{k^4} \tag{3.202}$$

Going to Euclidian space ($ik_0 = k_4$) we may evaluate this as

$$C_{\mu\nu\lambda\rho} = \frac{4i}{(2\pi)^2} \epsilon_{\mu\nu\lambda\epsilon} \int d^4k \frac{\partial}{\partial k_\rho} \frac{k_\epsilon}{k^4} \tag{3.203}$$

$$= \frac{4i}{(2\pi)^2} \epsilon_{\mu\nu\lambda\rho} \int d^4k \frac{\partial}{\partial k_\rho} \frac{k_\alpha}{4k^4} \tag{3.204}$$

where we used the fact that terms $\rho \neq \epsilon$ in Eq. (3.203) vanish by oddness in $k$; otherwise, $k_\alpha^2 = \frac{1}{4}k^2$ ($\alpha = 1, 2, 3, 4$). Now use Gauss's theorem to obtain

$$C_{\mu\nu\lambda\rho} = \frac{4}{(2\pi)^4} \epsilon_{\mu\nu\lambda\rho} \lim_{k\to\infty} \left(2\pi^2 k^3 \frac{k}{4k^4}\right) \tag{3.205}$$

$$= \frac{1}{8\pi^2} \epsilon_{\mu\nu\lambda\rho} \tag{3.206}$$

Using this result, we see that the shifts of integration variable $k' = (k + p_2)$ and $k'' = (k - p_1)$, respectively, give rise to

$$p_{1\mu} S_{\mu\nu\lambda} = \frac{1}{8\pi^2} \epsilon_{\mu\nu\lambda\rho} p_{2\rho} p_{1\mu} \tag{3.207}$$

$$p_{2\nu} S_{\mu\nu\lambda} = \frac{1}{8\pi^2} \epsilon_{\mu\nu\lambda\rho} p_{2\nu} p_{1\rho} \tag{3.208}$$

To include the crossed diagram (Fig. 3.7*b*) and to guarantee the vector Ward identities, we must use

$$\mathscr{S}_{\mu\nu\lambda}(p_1, p_2) = S_{\mu\nu\lambda}(p_1, p_2) + S_{\nu\mu\lambda}(p_2, p_1) + \frac{1}{4\pi^2} \epsilon_{\mu\nu\lambda\rho}(p_1 - p_2)_\rho \tag{3.209}$$

This satisfies

$$p_{1\mu} \mathscr{S}_{\mu\nu\lambda} = 0 \tag{3.210}$$

$$p_{2\nu} \mathscr{S}_{\mu\nu\lambda} = 0 \tag{3.211}$$

but also

$$(p_1 + p_2)_\lambda \mathscr{S}_{\mu\nu\lambda} = \frac{1}{2\pi^2} \epsilon_{\mu\nu\lambda\rho} p_{2\lambda} p_{1\rho} \tag{3.212}$$

Thus the axial-vector Ward identity contains an anomaly given uniquely by Eq. (3.212) and independent of the fermion mass in the triangle. No method of regularization can avoid this, for only one fermion is circulating in the triangle loop.

To make the point again, slightly differently, if we calculate the same Feynman diagram using different momentum labeling as in Fig. 3.8, we will arrive at amplitudes differing by a finite amount, namely,

$$T'_{\mu\nu\lambda} = T_{\mu\nu\lambda} + C_{\mu\nu\lambda\rho} p_{2\rho} \tag{3.213}$$

$$T''_{\mu\nu\lambda} = T_{\mu\nu\lambda} - C_{\mu\nu\lambda\rho} p_{1\rho} \tag{3.214}$$

with $C_{\mu\nu\lambda\rho}$ given by Eq. (3.206).

The presence of this anomaly helps reconcile the problem of the rate for $\lambda^0 \to \gamma\gamma$, which vanishes, for the unphysical limit $M_\pi \to 0$, according to the Sutherland–Veltman theorem [44, 45, 60].

Consider the decay

$$\pi^0(k) \to \gamma(p_1) + \gamma(p_2) \tag{3.215}$$

with matrix element

$$M(p_1, p_2) = \epsilon_\mu(p_1)\epsilon_\nu(p_2) T_{\mu\nu} \tag{3.216}$$

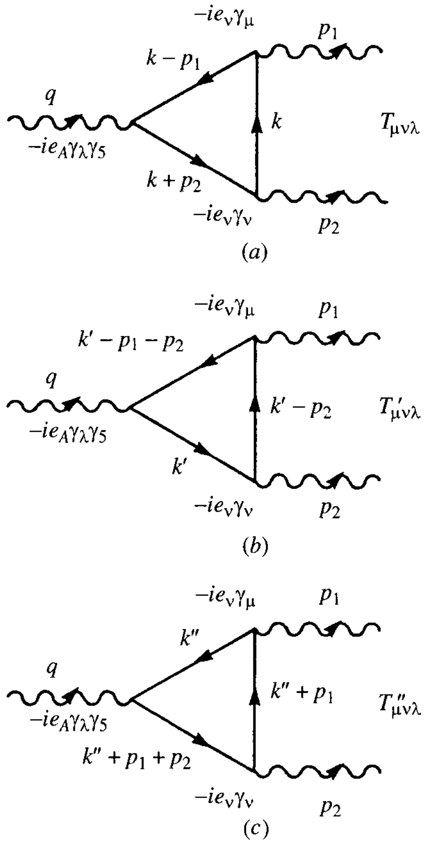



**Figure 3.8** Three different momentum labelings for the triangle graph amplitudes: (a) $T_{\mu\nu\lambda}$; (b) $T'_{\mu\nu\lambda}$; (c) $T''_{\mu\nu\lambda}$.

Then, by Lorentz invariance and parity conservation,

$$T_{\mu\nu} = \epsilon_{\mu\nu\alpha\beta} p_{1\alpha} p_{2\beta} T\left(k^2\right) \tag{3.217}$$

where the physical value is $k^2 = m_\pi^2$ but we consider $k^2$ a variable.

The PCAC hypothesis relates the pion field $\phi^a(x)$ to

$$\phi^a \sim \partial_\mu A_\mu^a \tag{3.218}$$

with proportionality constant $f_\pi/m_\pi^2$, $f_\pi$ = charged pion decay constant ($f_\pi \sim 0.96 m_\pi^3$). Then

$$T^{\mu\nu} \sim \left(m_\pi^2 - k^2\right) \int d^4x\, d^4y e^{-p_1\cdot x - p_2\cdot y} \partial_\lambda \langle 0|J^\mu(x) J^\nu(y) J^{5\lambda}(0)|0\rangle \tag{3.219}$$

$$= \left(m_\pi^2 - k^2\right) k_\lambda T_{\lambda\mu\nu} \tag{3.220}$$

where

$$T_{\lambda\mu\nu} \sim \int d^4x\, d^4y e^{-ip_1\cdot x - ip_2\cdot y} \langle 0|T\big(J^\mu(x)J^\nu(y)J^{5\lambda}(0)\big)|0\rangle \tag{3.221}$$

Since $T_{\lambda\mu\nu}$ is a pseudotensor, its most general invariant decomposition consistent with Bose statistics is

$$\begin{aligned} T_{\lambda\mu\nu} &= \epsilon^{\mu\nu\omega\phi} p_{1\omega} p_{2\phi} k_\lambda F_1\big(k^2\big) \\ &\quad + \big(\epsilon^{\lambda\nu\omega\phi} p_2^\nu - \epsilon^{\lambda\nu\omega\phi} p_1^\mu\big) p_{1\omega} p_{2\phi} F_2\big(k^2\big) \\ &\quad + \big(\epsilon^{\lambda\mu\omega\phi} p_1^\nu - \epsilon^{\lambda\nu\omega\phi} p_2^\mu\big) p_{1\omega} p_{2\phi} F_3\big(k^2\big) \\ &\quad + \epsilon^{\lambda\mu\nu\omega} (p_1 - p_2)_\omega \frac{1}{2} k^2 F_3\big(k^2\big) \end{aligned} \tag{3.222}$$

where the $F_i(k^2)$ have no kinematical singularities. [To derive Eq. (3.222), note that we require that $p_{1\mu} T_{\lambda\mu\nu} = P_{2\nu} T_{\lambda\mu\nu} = 0$ and $p_1^2 = p_2^2 = 0$, so that $p_1 \cdot p_2 = \frac{1}{2}k^2$].

But now we find immediately that

$$k_\lambda T^{\lambda\mu\nu} = \epsilon^{\mu\nu\omega\phi} p_1^\omega p_2^\phi k^2 \big[F_1\big(k^2\big) + F_3\big(k^2\big)\big] \tag{3.223}$$

and hence, in Eq. (3.217),

$$T\big(k^2 = 0\big) = 0 \tag{3.224}$$

Thus the rate $\pi^2 \to 2\gamma$ vanishes for $m_\pi^2 \to 0$ and is predicted to be very small for physical $m_\pi$, in contradiction to experiment.

The anomaly in Eq. (3.212) rescues this situation since Eq. (3.224) is no longer valid. Taking the anomaly into account, one arrives at a decay width

$$\Gamma = \frac{m_\pi^3}{64\pi} \left( \frac{2\sqrt{2} m_\pi^2 S}{\pi f_\pi} \right)^2 \alpha^2 \tag{3.225}$$

$$= \frac{m_\pi \alpha^2 S^2}{8\pi^3 (f_\pi / m_\pi^3)^2} \tag{3.226}$$

$$= (7.862\ \text{eV})(2S)^2 \tag{3.227}$$

where $S$ is given by the trace over the $\gamma_5$ couplings for the fermions circulating in the triangle. With three colored up and down quarks participating in the pion function, one has

$$S = \text{Tr}\big(g^5 3 Q^2\big) \tag{3.228}$$

$$= 3 \times \frac{1}{2}\big(e_\mu^2 - e_d^2\big) \tag{3.229}$$

$$= \frac{3}{2}\left(\frac{4}{9} - \frac{1}{9}\right) \tag{3.230}$$

$$= \frac{1}{2} \tag{3.231}$$

Then Eq. (3.227) agrees beautifully with the observed width [61]

$$\Gamma\left(\pi^0 \to 2\gamma\right) = 7.8 \pm 0.9 \text{ eV} \tag{3.232}$$

In the absence of the color degree of freedom, the rate is too small by a factor of 9. Thus this calculation both supports the existence of the triangle anomaly and provides support for the color idea. It has been assumed, however, that nonperturbative effects associated with the quark–antiquark confinement in the pion can be neglected.

To allow renormalization such that the renormalized Lagrangian $\mathscr{L}_R$ in Eq. (3.1) be gauge invariant, the anomaly in the axial Ward identity must be canceled between different fermion species (recall that the anomaly is mass independent). Let us focus on an SU(2) × U(1) theory with gauge bosons $A_\mu^a$ ($a = 1, 2, 3$) and $B_\mu$. Let the fermions have electric charges $Q_i$ given by

$$Q_i = I_i^3 + \frac{1}{2} Y_i \tag{3.233}$$

where $I^3$ is the third component of weak isospin and $Y$ is the weak hypercharge.

We need not consider triangles of the type $(A^a A^b A^c)$ or $(A^a B B)$ because these involve the trace of an odd number of isospin matrices and hence vanish. We need look only at $(BBB)$, $(A^0 A^0 B)$, and $(A^+ A^- B)$. Let us distinguish left and right helicities by subscripts $L$ and $R$, respectively; these correspond to the decomposition

$$\psi_i = \frac{1}{2}(1+\gamma_5)\psi_i + \frac{1}{2}(1-\gamma_5)\psi_i \tag{3.234}$$

$$= \psi_{iR} + \psi_{iL} \tag{3.235}$$

Clearly, $L$ and $R$ have opposite sign $\gamma_5$ couplings, and since we are always focusing on the triangle with one axial and two vector couplings (AVV in the notation of Fig. 3.6) the cancellation of the $BBB$ anomaly requires that

$$\sum_{i_L} Y_{i_L}^3 = \sum_{i_R} Y_{i_R}^3 \tag{3.236}$$

We may rewrite this

$$\sum_{i_L} \left(Q_{i_L} - I_{i_L}^3\right)^3 = \sum_{i_R} \left(Q_{i_R} - I_{i_R}^3\right)^3 \tag{3.237}$$

Consider any given multiplet, either left-handed $\{i_L\} \subset M_L$ or right-handed $\{i_R\} \subset M_R$; then within a multiplet one has

$$\sum_{i \subset M} \left(I_i^3\right)^3 = 0. \tag{3.238}$$

Given that no massless charged fermions occur, then

$$\sum_{i_L} Q_{i_L}^3 = \sum_{i_R} Q_{i_R}^3 \tag{3.239}$$

Hence, from Eq. (3.237),

$$\sum_{i_L} \left[Q_{i_L}^2 \left(I_{i_L}^3\right) + Q_{i_L} \left(I_{i_L}^3\right)^2\right] = \sum_{i_R} \left[Q_{i_R}^2 \left(I_{i_R}^3\right) + Q_{i_R} \left(I_{i_R}^3\right)^2\right] \tag{3.240}$$

Now consider the flavor factors for $(A^+ A^- B)$ and use the fact that

$$2I^+ I^- = (I)^2 - \left(I^3\right)^2 \tag{3.241}$$

to arrive at the condition

$$\sum_{i_L} Y_{i_L} \left[(I_{i_L})^2 - \left(I_{i_L}^3\right)^2\right] = \sum_{i_R} Y_{i_R} \left[(I_{i_R})^2 - \left(I_{i_R}^3\right)^2\right] \tag{3.242}$$

Consider the factors involved in $(BA^0 A^0)$ to find, in addition,

$$\sum_{i_L} Y_{i_L} \left(I_{i_L}^3\right)^2 = \sum_{i_R} Y_{i_R} \left(I_{i_R}^3\right)^2 \tag{3.243}$$

Using Eqs. (3.233) and (3.238), this gives

$$\sum_{i_L} Q_{i_L} \left(I_{i_L}^3\right)^2 = \sum_{i_R} Q_{i_R} \left(I_{i_R}^3\right)^2 \tag{3.244}$$

With Eq. (3.240), this implies that

$$\sum_{i_L} Q_{i_L}^2 \left(I_{i_L}^3\right) = \sum_{i_R} Q_{i_R}^2 \left(I_{i_R}^3\right) \tag{3.245}$$

Using Eq. (3.233) in Eq. (3.242) gives a third relation:

$$\sum_{i_L} Q_{i_L} \left(I_{i_L}\right)^2 = \sum_{i_R} Q_{i_R} \left(I_{i_R}\right)^2 \tag{3.246}$$

The three requirements, Eqs. (3.244) through (3.246), will now be shown to be equivalent. For consider within one multiplet the sum

$$\sum_{i \subset M} Q_i^2 (I_i^3) = \sum_{i \subset M} \left( \frac{Y_M}{2} + I_i^3 \right)^2 I_i^3 \tag{3.247}$$

$$= Y_M \sum_{i \subset M} (I_i^3)^2 \tag{3.248}$$

$$= 2 \sum_i (Q_i - I_i^3)(I_i^3)^2 \tag{3.249}$$

$$= \sum_{i \subset M} Q_i (I_i^3)^2 \tag{3.250}$$

showing that Eqs. (3.244) and (3.245) are equivalent.

Now

$$\sum_{i \subset M} (I_i^3)^2 = 2 \sum_{M=1}^{I_M} M^2 \tag{3.251}$$

$$= \frac{1}{3} I_M (I_M + 1)(2I_M + 1) \tag{3.252}$$

valid for $I_M =$ integer or half-integer (where $M$ sum starts with $M + \frac{1}{2}$).

Also,

$$\sum_{i \subset M} Q_i = \sum_{i \subset M} \left( I_i^3 + \frac{1}{2} Y_M \right) \tag{3.253}$$

$$= \frac{1}{2} Y_M (2I_M + 1) \tag{3.254}$$

Hence, for any complete representation,

$$\sum_{i \subset M} (Q_i)^2 (I_i^3) = 2 \sum_{i \subset M} Q_i (I_i^3)^2 \tag{3.255}$$

$$= \frac{2}{3} \sum_{i \subset M} (I_M)^2 Q_i \tag{3.256}$$

and thus Eq. (3.246) is also not independent.

The anomaly-free condition for SU(2) × U(1) quantum flavor dynamics is therefore summarized by

$$\sum_{M_L} I_{M_L} (I_{M_L} + 1) \sum_{i_L \subset M_L} Q_{i_L} = \sum_{M_R} I_{M_R} (I_{M_R} + 1) \sum_{i_R \subset M_R} Q_{i_R} \tag{3.257}$$

Thus $I = 0$ does not contribute. In a theory with only singlets and left-handed doublets, the condition reduces to

$$\sum_{i_L} Q_{i_L} = 0 \tag{3.258}$$

In a model with three colors of quark doublets with charges $(\frac{2}{3}, -\frac{1}{3})$ and an equal number of lepton doublets with charges $(0, -1)$, the anomaly cancels between quarks and leptons.

## 3.4
## Becchi–Rouet–Stora–Tyutin Transformation

To ensure that renormalization is consistent with local gauge invariance, as is essential to preserve perturbative unitarity, the remarkable BRST identities [37] (and anti-BRST identities [62–66]) are much simpler to implement than the Taylor–Slavnov [38, 39] identities to which they are equivalent.

Let us recall how classical gauge invariance works to establish notation. From Chapter 1 we have

$$\delta A^i_\mu = -\frac{1}{g}\partial_\mu \theta^i + c_{ijk}\theta^j A^k_\mu \tag{3.259}$$

$$= -\frac{1}{g}(D_\mu \theta)^i \tag{3.260}$$

for an infinitesimal gauge function $\boldsymbol{\theta}(x)$. Defining

$$c_{ijk} A_j B_k = (\mathbf{A} \hat{} \mathbf{B})_i \tag{3.261}$$

we have

$$\delta \mathbf{A}_\mu = -\frac{1}{g}\partial_\mu \boldsymbol{\theta} + \boldsymbol{\theta} \hat{} \mathbf{A}_\mu \tag{3.262}$$

$$\delta \mathbf{F}_{\mu\nu} = \boldsymbol{\theta} \hat{} \mathbf{F}_{\mu\nu} \tag{3.263}$$

Since $\delta \mathbf{F}_{\mu\nu}$ is perpendicular to $\mathbf{F}_{\mu\nu}$, we see that $\mathscr{L}_{\text{cl}} = -\frac{1}{4}(\mathbf{F}_{\mu\nu} \cdot \mathbf{F}_{\mu\nu})$ is invariant.

Upon quantization of the theory, in Chapter 2 we arrived at an effective Lagrangian,

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_{\text{cl}} + \mathscr{L}_{\text{GF}} + \mathscr{L}_{\text{FPG}} \tag{3.264}$$

where the gauge-fixing term is, for example,

$$\mathscr{L}_{\text{GF}} = -\frac{1}{2\alpha}(\partial_\mu \mathbf{A}_\mu)^2 \tag{3.265}$$

for the covariant gauges. (The discussion of BRST transformations that follows is easily modified for other choices of gauge.) The Faddeev–Popov ghost term is

$$\mathscr{L}_{\text{FPG}} = \partial_\mu \mathbf{c}^+ \cdot D_\mu \mathbf{c} \tag{3.266}$$

where $\mathbf{c}(x)$ and $\mathbf{c}^+(x)$ are the anticommuting scalar ghost fields.

The ingenious idea of Becchi et al. [37] is to relate the gauge function to the ghost field $\mathbf{c}(x)$ by

$$\boldsymbol{\theta}(x) = g\mathbf{c}(x)\delta\lambda \tag{3.267}$$

where $\delta\lambda$ is an (infinitesimal) anticommuting $c$-number. We also define the operation

$$\frac{d}{d\lambda} \tag{3.268}$$

with the understanding that $\delta\lambda$ has been placed at the extreme right of the expression acted upon.

The BRS transformation is now defined by

$$\delta \mathbf{A}_\mu = -(D_\mu \mathbf{c})\delta\lambda \tag{3.269}$$

$$\delta \mathbf{c} = -\frac{1}{2} g(\mathbf{c}\,\hat{}\,\mathbf{c})\delta\lambda \tag{3.270}$$

$$\delta \mathbf{c}^+ = -\frac{1}{\alpha}(\partial_\mu \mathbf{A}_\mu)\delta\lambda \tag{3.271}$$

We must first show that $\mathscr{L}_{\text{eff}}$ in Eq. (3.264) is invariant under these transformations. That $\mathscr{L}_{\text{cl}}$ is invariant follows trivially, since we have only reparametrized $\boldsymbol{\theta}(x)$. For $\mathscr{L}_{\text{GF}}$ and $\mathscr{L}_{\text{FPG}}$ we find that

$$\delta \mathscr{L}_{\text{GF}} = \frac{1}{\alpha}(\partial_\lambda \mathbf{A}_\lambda)(\partial_\mu D_\mu \mathbf{c})\delta\lambda \tag{3.272}$$

$$\begin{aligned}\delta \mathscr{L}_{\text{FPG}} = {} & -\frac{1}{\alpha}\left[\partial_\mu(\partial_\lambda \mathbf{A}_\lambda)\right]\delta\lambda(\partial_\mu \mathbf{c} + g\mathbf{A}_\mu\,\hat{}\,\mathbf{c}) \\ & + \partial_\mu \mathbf{c}^+ \cdot \left[-\frac{1}{2} g\partial_\mu(\mathbf{c}\,\hat{}\,\mathbf{c})\partial\lambda - gD_\mu \mathbf{c}\delta\lambda\,\hat{}\,\mathbf{c} - \frac{g^2}{2}\mathbf{A}_\mu\,\hat{}\,(\mathbf{c}\,\hat{}\,\mathbf{c})\delta\lambda\right]\end{aligned} \tag{3.273}$$

Bearing in mind that $\delta\lambda$ anticommutes with $\mathbf{c}$ and $\mathbf{c}^+$, we have

$$\frac{d}{d\lambda}(\mathscr{L}_{\mathrm{GF}} + \mathscr{L}_{\mathrm{FPG}}) = \frac{1}{\alpha}\partial_\mu\left[(\partial_\mu \mathbf{A}_\lambda)\cdot(D_\mu \mathbf{c})\right]$$

$$+ \partial_\mu \mathbf{c}^+\left[-\frac{1}{2}g\partial_\mu(\mathbf{c}\hat{}\mathbf{c}) + g(D_\mu\mathbf{c})\hat{}\mathbf{c}\right.$$

$$\left. - \frac{1}{2}g^2\mathbf{A}_\mu\hat{}(\mathbf{c}\hat{}\mathbf{c})\right] \tag{3.274}$$

$$\equiv 0 \tag{3.275}$$

The first term in Eq. (3.274) vanishes in the variation of the action, since it is a total derivative. That the second term vanishes follows from the relations

$$\partial_\mu(\mathbf{c}\hat{}\mathbf{c}) = 2(\partial_\mu\mathbf{c})\hat{}\mathbf{c} \tag{3.276}$$

$$A_\mu\hat{}(\mathbf{c}\hat{}\mathbf{c}) = 2(A_\mu\hat{}\mathbf{c})\hat{}\mathbf{c} \tag{3.277}$$

(To prove these, use $\mathbf{c}_1$ and $\mathbf{c}_2$, then put $\mathbf{c}_1 = \mathbf{c}_2 = \mathbf{c}$.) Thus

$$\int d^4x\, \mathscr{L}_{\mathrm{eff}} \tag{3.278}$$

is invariant under the BRST transformation.

To show invariance of the generating functional and hence of the Green's functions, we need to consider the functional Jacobian or determinant

$$\frac{A_\mu^a(x) + \delta A_\mu^a(x),\, c^a(x) + \delta c^a(x),\, c^{+a}(x) + \delta c^{+a}(x)}{A_\nu^b(y),\, c^b(y),\, c^{+b}(y)} \tag{3.279}$$

The only nonvanishing elements in the determinant are

$$\frac{\delta[A_\mu^a(x) + \delta A_\mu^a(x)]}{\delta A_\nu^b(y)} = g_{\mu\nu}\delta^4(x-y)\left(\delta^{ab} - gf^{abc}c^c\delta\lambda\right) \tag{3.280}$$

$$\frac{\delta(c^a(x) + \delta c^a(x))}{\delta c^b(y)} = \delta^4(x-y)\left(\delta^{ab} + gf^{abc}c^c\delta\lambda\right) \tag{3.281}$$

$$\frac{\delta(A_\mu^a(x) + \delta A_\mu^a(x))}{\delta c^b(y)} = \delta^4(x-y)\left(gf^{abc}A_\mu^c\delta\lambda\right) \tag{3.282}$$

Only the second of these is tricky: One must put the relevant $\mathbf{c}$ to the extreme right before differentiating, as follows:

$$\frac{\delta}{\delta c^b(y)}\left[\delta c^a(x)\right] = -\frac{g}{2}\frac{\delta}{\delta c^b(y)}\left[f^{acd}c^c(x)c^d(x)\right]\delta\lambda \tag{3.283}$$

$$= -\frac{g}{2}\delta^4(x-y)\left[f^{acd}c^c(x)\delta^{db}\right.$$

$$- f^{acd} c^d(x)\delta^{cb}\big]\delta\lambda \tag{3.284}$$

$$= +g\delta^4(x-y) f^{abc} c^c(x)\delta\lambda \tag{3.285}$$

as required.

The relevant determinant, in block form, is now

$$\left| \begin{pmatrix} 1 - gfc\delta\lambda & gfA\delta\lambda & 0 \\ 0 & 1 + gfc\delta\lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \right| \left[\delta^4(x-y)\right]^3 g_{\mu\nu} \tag{3.286}$$

and since $(\delta\lambda)^2 = 0$, we are left with the unit matrix.

Thus

$$\frac{d}{d\lambda} W[J] = 0 \tag{3.287}$$

and all Green's functions have global BRST invariance. This simplifies and systematizes the original Taylor–Slavnov form of the Ward identities.

It is useful to note the following nilpotencies:

$$\delta \mathbf{A}_\mu = -D_\mu \mathbf{c}\delta\lambda \tag{3.288}$$

$$\delta^2 \mathbf{A}_\mu = \frac{g}{2}\partial_\mu(\mathbf{c}\,\hat{}\,\mathbf{c})\delta\lambda_2\delta\lambda_1 + gD_\mu \mathbf{c}\delta\lambda_2\,\hat{}\,\mathbf{c}\delta\lambda_1$$

$$+ \frac{g^2}{2}\mathbf{A}_\mu\,\hat{}\,(\mathbf{c}\,\hat{}\,\mathbf{c})\delta\lambda_2\delta\lambda_1 \tag{3.289}$$

$$= 0 \tag{3.290}$$

where we used Eqs. (3.276) and (3.277). Also

$$\delta \mathbf{c} = -\frac{q}{2}\mathbf{c}\,\hat{}\,\mathbf{c}\delta\lambda \tag{3.291}$$

$$\delta^2 c = \frac{1}{4}g^2\left[(\mathbf{c}\,\hat{}\,\mathbf{c})\delta\lambda_2\,\hat{}\,\mathbf{c}\delta\lambda_1 + \mathbf{c}\,\hat{}\,(\mathbf{c}\,\hat{}\,\mathbf{c})\delta\lambda_2\delta\lambda_1\right] \tag{3.292}$$

$$= 0 \tag{3.293}$$

Finally, one has

$$\delta \mathbf{c}^+ = \frac{-1}{\alpha}(\partial_\mu \mathbf{A}_\mu)\delta\lambda \tag{3.294}$$

and hence, from Eq. (3.271),

$$\delta^2 \mathbf{c}^+ = -\frac{1}{\alpha}\left(\partial_\mu \delta^2 \mathbf{A}_\mu\right)\delta\lambda \tag{3.295}$$

$$= 0 \tag{3.296}$$

Following from the BRST invariance of $\mathscr{L}_{\text{eff}}$, as demonstrated above, we can straightforwardly construct the corresponding BRST Noether current. It is given by

$$J_\mu^{\text{BRST}} = \frac{\partial \mathscr{L}_{\text{eff}}}{\partial(\partial_\mu \phi_k)}(\delta\phi_k)_{\text{BRST}} \tag{3.297}$$

where the sum in $k$ is over all participating fields. Now

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_{\text{cl}} + \mathscr{L}' \tag{3.298}$$

and

$$\frac{\partial \mathscr{L}_{\text{cl}}}{\partial(\partial_\mu A_\nu^a)} = -F_{\mu\nu}^a \tag{3.299}$$

$$\frac{\partial \mathscr{L}'}{\partial(\partial_\mu A_\nu^a)} = -\frac{1}{\alpha} g_{\mu\nu}\left(\partial_\lambda A_\lambda^a\right) \tag{3.300}$$

$$\frac{\partial \mathscr{L}'}{\partial(\partial_\mu c^a)} = \partial_\mu c^{+a} \tag{3.301}$$

$$\frac{\partial \mathscr{L}'}{\partial(\partial_\mu c^{+a})} = (D_\mu \mathbf{c})_a \tag{3.302}$$

whence

$$J_\mu^{\text{BRS}} = \left[-\mathbf{F}_{\mu\nu} D_\nu \mathbf{c} - \frac{g}{2}\partial_\mu \mathbf{c}^+ \cdot (\mathbf{c} \,\hat{}\, \mathbf{c})\right]\delta\lambda \tag{3.303}$$

is the required conserved current. We have used the fact that

$$\frac{\partial \mathscr{L}'}{\partial(\partial_\mu A_\nu^a)}\left(\delta A_\nu^a\right)_{\text{BRST}} = -\frac{\partial \mathscr{L}}{\partial(\partial_\mu c^{+a})}\left(\delta c^{+a}\right)_{\text{BRST}} \tag{3.304}$$

## 3.5 Proof of Renormalizability

Historically, the renormalizability of Yang–Mills theory, for both unbroken and broken symmetry, was first made clear in the classic papers by 't Hooft [67, 68]. [For subsequent formal developments, see Refs. [11] and [69–81]; for specific examples of the procedure at low orders, see, e.g., Refs. [82–88]; for salient earlier work on renormalization of the (nongauge) sigma model, see Refs. [89–93]; for reviews, see Refs. [14] and [94–100].] This work confirmed the earlier conjecture by Weinberg [101] and Salam [102] that renormalizability of a gauge theory survived under spontaneous symmetry breakdown.

In our treatment here (see, e.g., Ref. [103]), we rely heavily on the BRST transformation introduced in Section 3.4.

Also, although its explicit use is avoided, the existence of a regularization scheme that maintains gauge invariance (i.e., dimensional regularization) is absolutely crucial. Thus, closed fermion loops involving $\gamma_5$ are still excluded.

We shall first exclude all matter fields and concentrate on pure Yang–Mills theory. Suppose that there are $E_v$ external gluons, the only possible external particle; then since each extra gluon adds extra coupling $g$, the power of $g(q_0)$ in a tree diagram is

$$q_0 = E_v - 2 \tag{3.305}$$

Note that we are interested only in connected diagrams (use $Z \sim \ln W$ as generating functional); however, we are not restricting attention to proper, or one-particle-irreducible (1PI) diagrams. Since each extra loop adds $g^2$, with $p$ loops we have a power ($q$) of $g$ given by

$$q = q_0 + 2p \tag{3.306}$$

The procedure will be iterative in $p$, the number of loops; that is, we assume the theory has been rendered finite for $(p-1)$ loops, and consider $p$ loops. Since $p = 1$ is trivially finite—tree diagrams have no momentum integrals—all $p$ will then follow. Because of Eq. (3.306), iteration in $p$ is precisely equivalent to iteration in powers of $g$, for fixed $E_v$.

We shall systematically assume that for any given diagram all subdivergences have been handled already, using known theorems. Such theorems are enshrined in the Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) renormalization procedure, for example, and involve subtle treatment of overlapping divergences, and so on. The interested reader who wishes to study the BPHZ program is referred to Lee's review [98, Sec. 3] and the references cited therein.

We begin by defining renormalization constants $z$, $z_3$, and $X$ according to

$$A_\mu^a = \sqrt{Z_3} A_{\mu R}^a \tag{3.307}$$

$$c^a = \sqrt{\mathscr{Z}} c_R^a \tag{3.308}$$

$$c^{+a} = \sqrt{\mathscr{Z}} c_R^{+a} \tag{3.309}$$

$$g = \frac{X}{\mathscr{Z}\sqrt{Z_3}} g_R \tag{3.310}$$

$$\delta\lambda = \sqrt{\mathscr{Z} Z_3}\, \delta\lambda_R \tag{3.311}$$

$$\alpha = Z_3 \alpha_R \tag{3.312}$$

It will be convenient to define the two auxiliary operators $\mathbf{K}_\mu$ and $\mathbf{L}$ according to

$$\mathbf{K}_\mu = D_\mu \mathbf{c} \tag{3.313}$$

$$\mathbf{L} = \mathbf{c} \hat{} \mathbf{c} \tag{3.314}$$

These operators are renormalized as follows:

$$\mathbf{K}_\mu = \frac{1}{\sqrt{\mathscr{X}}} \mathbf{K}_{\mu R} \tag{3.315}$$

$$\mathbf{L} = \mathscr{X} \mathbf{L}_R \tag{3.316}$$

Recall that the BRS transformations for unrenormalized quantities are given by

$$\delta \mathbf{A}_\mu = -\mathbf{K}_\mu \delta\lambda \tag{3.317}$$

$$\delta \mathbf{c} = -\frac{1}{2} g \mathbf{L} \delta\lambda \tag{3.318}$$

$$\delta \mathbf{c}^+ = -\frac{1}{\alpha} (\partial_\mu \mathbf{A}_\mu) \delta\lambda \tag{3.319}$$

Using Eqs. (3.307) through (3.316), the renormalized versions become

$$\delta \mathbf{A}_{\mu R} = -K_{\mu R} \delta\lambda_R \tag{3.320}$$

$$\delta \mathbf{c}_R = -\frac{1}{2} X g_R \mathbf{L}_R \delta\lambda_R \tag{3.321}$$

$$\delta \mathbf{c}_R^+ = -\frac{1}{\alpha_R} (\partial_\mu \mathbf{A}_{\mu R}) \delta\lambda_R \tag{3.322}$$

The task before us is to show that all Green's functions are finite and that the operators $\mathbf{K}_{\nu R}$ and $\mathbf{L}_R$ are finite operators (i.e., their vacuum expectation values with any combination of renormalized basic fields is finite).

Recall from Chapter 2 that the degree of superficial divergence is given by $D = 4 - E_B$, where $E_B$ is the number of external particles (recall that the vector states are massless).

The primitively divergent diagrams are shown in Fig. 3.9. For Fig. 3.9*a* and *b* the effective degree of divergence is reduced one power by the derivative always present on the outgoing ghost line. Note that the full vector propagator (Fig. 3.9*c*) has three contributions, as indicated in the figure, all of which must be considered separately. Also, remark that the two ghost–two antighost Green's function (Fig. 3.10), superficially logarithmically divergent, is rendered finite by the two external derivatives.

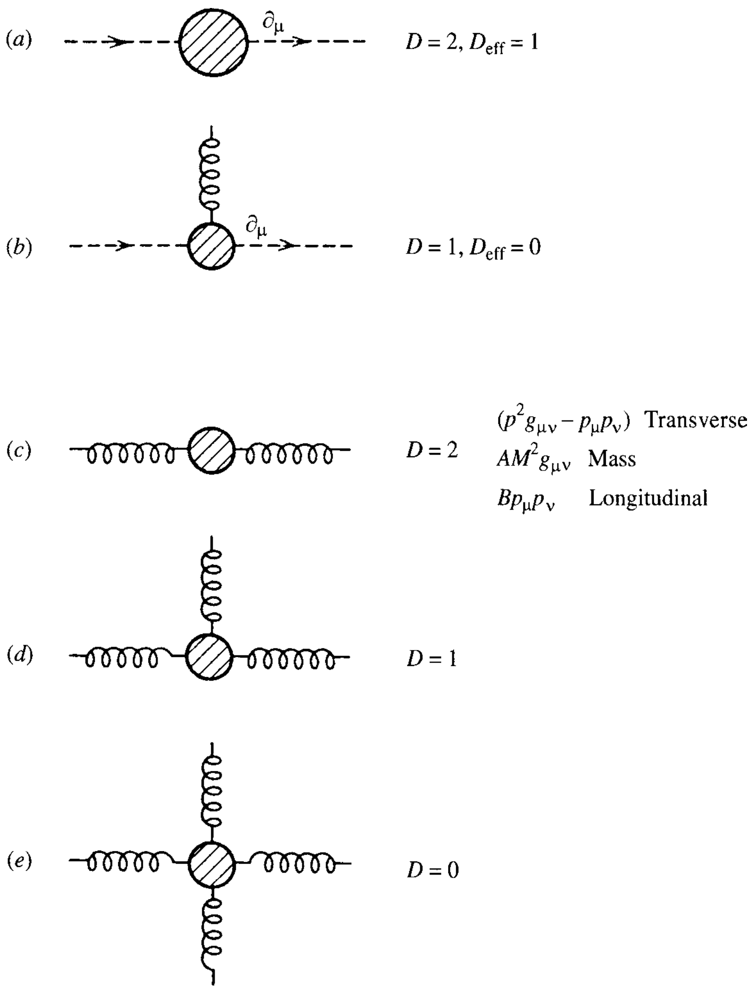



**Figure 3.9** Primitively divergent diagrams for Yang–Mills theory.

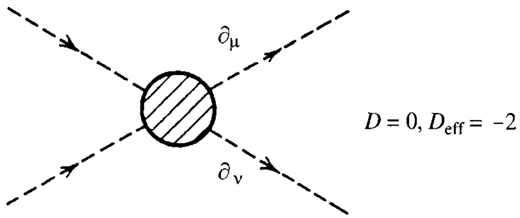



**Figure 3.10** Two ghost–two antighost Green's function.

At the $p$-loop level, we proceed by defining the renormalization constants $Z_3$, $\mathscr{L}$, and $X$ such that the transverse part of Fig. 3.9*c*, the Green's functions (Fig. 3.9*a* and *b*), are finite, respectively. Now we must demonstrate the five further quantities are finite: (1) $\mathbf{K}_{\mu R}$, (2) $\mathbf{L}_R$, (3) the mass and longitudinal parts of the two-vector diagram, (4) the triple-vector diagram, and finally, (5) the four-vector coupling. This we do in five steps.

**Step 1: $\mathbf{K}_{\mu R}$.** The unrenormalized effective Lagrangian, Eq. (3.264), depends on the ghost fields according to

$$\mathscr{L}_{\mathrm{FPG}} = \partial_\mu \mathbf{c}^+ \cdot D_\mu \mathbf{c} \tag{3.323}$$

In the renormalized Lagrangian there appears the term

$$\partial_\mu \mathbf{c}_R^+ \cdot D_\mu \mathbf{c}_R = -\mathbf{c}_R^+ \cdot \partial_\mu D_\mu \mathbf{c}_R + \text{total derivative} \tag{3.324}$$

Thus the renormalized equation of motion for the ghost field

$$\frac{\partial \mathscr{L}_R^{\mathrm{eff}}}{\partial \mathbf{c}_R^+} = \partial_\mu \frac{\partial \mathscr{L}_R^{\mathrm{eff}}}{\partial(\partial_\mu \mathbf{c}_R^+)} \tag{3.325}$$

becomes

$$\frac{\partial \mathscr{L}_R^{\mathrm{eff}}}{\partial \mathbf{c}_R^+} = -\partial_\mu D_\mu \mathbf{c}_R \tag{3.326}$$

The way in which to exploit this equation of motion is to consider the vacuum expectation of the time-ordered product of either side of Eq. (3.326) with any combination of renormalized fundamental fields (hereafter denoted by "fields"). Then one can derive that

$$\langle 0|T\big[(\partial_\mu D_\mu \mathbf{c}_R)(\text{“fields”})\big]|0\rangle = -i\langle 0|T\left[\frac{\partial}{\partial \mathbf{c}_R^+}(\text{“fields”})\right]|0\rangle \tag{3.327}$$

This is derived by noting that the left-hand side of Eq. (3.327) is given by

$$-\int D\mathbf{A}_\mu D\mathbf{c} D\mathbf{c}^+ \frac{\partial \mathscr{L}}{\partial \mathbf{c}_R^+}(\text{“fields”})e^{iS_{\mathrm{eff}}}$$

$$= +i\int D\mathbf{A}_\mu D\mathbf{c} D\mathbf{c}^+ (\text{“fields”})\frac{\partial}{\partial \mathbf{c}_R^+}\left(e^{iS_{\mathrm{eff}}}\right) \tag{3.328}$$

$$= -i\int D\mathbf{A}_\mu D\mathbf{c} D\mathbf{c}^+ \left[\frac{\partial}{\partial \mathbf{c}_R^+}(\text{“field”})\right]e^{iS_{\mathrm{eff}}} \tag{3.329}$$

which equals the right-hand side of Eq. (3.327). In the last step, Eq. (3.329), we have used integration by parts and dropped a surface term.

Now what combinations of gauge and ghost fields can make the left-hand side Eq. (3.327) divergent? Only two possibilities exist: $\mathbf{c}_R^+$ and $\mathbf{c}_R^+\mathbf{A}_{\mu R}$. But on the right-hand side, one has, respectively,

$$\langle 0|1|0\rangle = 1 < \infty \tag{3.330}$$

$$\langle 0|\mathbf{A}_\mu|0\rangle = 0 \tag{3.331}$$

Thus $\mathbf{K}_{\mu R}$ is a finite operator, as required.

**Step 2: ($\mathbf{L}_R$).** Consider the fact that

$$\frac{d}{d\lambda_R}\langle 0|T\left(c_R c_R^+ c_R^+\right)|0\rangle = 0 \tag{3.332}$$

This equation is written in a shorthand that will be used repeatedly. More explicitly, using the Noether current $J_\mu^{\mathrm{BRST}}$ of Section 3.4, what we mean is

$$-\int d^4x\,\langle 0|T\left(\partial_\mu J_\mu^{\mathrm{BRST}}(x) c_R c_R^+ c_R^+\right)|0\rangle = 0 \tag{3.333}$$

Integration by parts and dropping surface terms gives Eq. (3.332).
Expanding Eq. (3.332) then gives

$$\langle 0|\delta c_R c_R^+ c_R^+|0\rangle + \langle 0|c_R \delta c_R^+ c_R^+|0\rangle + \langle 0|c_R c_R^+ \delta c_R^+|0\rangle = 0 \tag{3.334}$$

The last two terms are finite by Eq. (3.322) and the definition of $X$. Hence $L_R$ is a finite operator.

**Step 3: Gluon Propagator.** Consider

$$0 = \frac{d}{d\lambda_R}\langle 0|T\left(c_R^+ A_{\mu R}\right)|0\rangle \tag{3.335}$$

$$= \langle 0|\delta c_R^+ A_{\mu R}|0\rangle + \langle 0|c_R^+ \delta A_{\mu R}|0\rangle \tag{3.336}$$

$$= \frac{1}{\alpha_R}\langle 0|\partial_\lambda A_{\lambda R} A_{\mu R}|0\rangle + \langle 0|c_R^+ \delta A_{\mu R}|0\rangle \tag{3.337}$$

The second term of Eq. (3.337) is already known to be finite, $\alpha_R$ is chosen to be finite as a gauge choice, and hence the mass and longitudinal parts of Fig. 3.9*c* are finite.

**Step 4: Three-Gluon Vertex.** Consider

$$0 = \frac{d}{d\lambda_R}\langle 0|T\left(c_R^+ A_{\mu R} A_{\nu R}\right)|0\rangle \tag{3.338}$$

$$= \langle 0|\delta c_R^+ A_{\mu R} A_{\nu R}|0\rangle + \langle 0|c_R^+ \delta A_{\mu R} A_{\nu R}|0\rangle + \langle 0|c_R^+ A_{\mu R} \delta A_{\nu R}|0\rangle \tag{3.339}$$

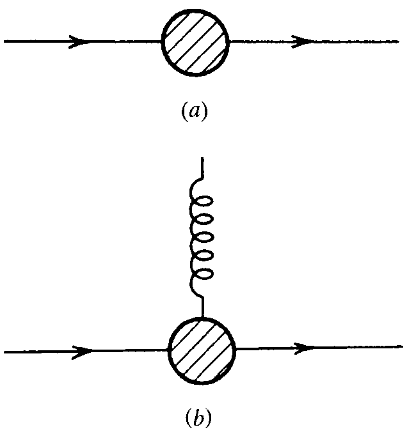



**Figure 3.11** Extra divergences when fermions are added.

The second and third terms are finite, as, therefore, is

$$\frac{1}{\alpha_R}\partial_\lambda \langle 0|A_{\lambda R}A_{\mu R}A_{\nu R}|0\rangle \tag{3.340}$$

as required.

**Step 5: Four-Gluon Vertex.** Consider

$$0 = \frac{d}{d\lambda_R}\langle 0|T\left(c_R^{+}A_{\lambda R}A_{\mu R}A_{\nu R}\right)|0\rangle \tag{3.341}$$

$$= \frac{-\partial_\kappa}{\alpha_R}\langle 0|A_{\kappa R}A_{\lambda R}A_{\mu R}A_{\nu R}|0\rangle + \text{finite terms involving } \mathbf{K}_{\mu R} \tag{3.342}$$

Thus the four-gluon connected Green's function is finite. This completes the proof for pure Yang–Mills theory.

Now we must add matter fields: first, spin-$\frac{1}{2}$ fermions (as in quantum chromodynamics, for example). Allowing the fermions to fall in an arbitrary representation of the gauge group, their BRST transformation is

$$\delta\psi = -g\mathbf{T}\cdot\mathbf{c}\psi\delta\lambda \tag{3.343}$$

Here $\mathbf{T}$ is a vector of the same dimension as $\mathbf{A}_\mu$; it is also a matrix with the same dimension as $\psi$. We introduce two renormalization constants, $Z_2$ and $Z_m^f$, by

$$\psi = \sqrt{Z_2}\psi_R \tag{3.344}$$

$$m_f = \sqrt{Z_m^f}m_{fR} \tag{3.345}$$

These are defined such that the full fermion propagator (Fig. 3.11*d*) is finite (both the $A\not{p}$ and the $Bm$ pieces).

Now consider the renormalized BRST transformation

$$\delta\psi_R = Xg_R\mathbf{T}\cdot\mathbf{c}_R\psi_R\delta\lambda_R \tag{3.346}$$

To prove that $d\psi_R/d\lambda_R$ is finite involves one tricky argument. Consider the ghost term in the renormalized Lagrangian

$$\partial_\mu\mathbf{c}_R^+(\mathscr{Z}\partial_\mu\mathbf{c}_R + Xg_R\mathbf{A}_{\mu R}\hat{}\mathbf{c}_R) = \left(-\mathscr{Z}\,\Box\,\mathbf{c}_R^+ + Xg_R\partial_\mu\mathbf{c}_R^+\hat{}\mathbf{A}_{\mu R}\right)\cdot\mathbf{c}_R \tag{3.347}$$

where we discard total differentials. Thus the ghost equation of motion involves

$$\frac{\partial\mathscr{L}_R^{\text{eff}}}{\partial\mathbf{c}_R} = -\mathscr{Z}\,\Box\,\mathbf{c}_R^+ + Xg_R\partial_\mu\mathbf{c}_R^+\hat{}\mathbf{A}_{\mu R} \tag{3.348}$$

Hence, using the same steps as for Eq. (3.327), we have

$$\langle 0|\left(-\mathscr{Z}\,\Box\,\mathbf{c}_R^+ + Xg_R\partial_\mu\mathbf{c}_R^+\hat{}\mathbf{A}_{\mu R}\right)\delta\psi_R\bar{\psi}_R|0\rangle$$

$$= +i\left\langle 0\left|\frac{\delta}{\delta\mathbf{c}_R}(\delta\psi_R\bar{\psi}_R)\right|0\right\rangle \tag{3.349}$$

$$= -iXg_R\mathbf{T}\langle 0|T(\psi_R\bar{\psi}_R)|0\rangle \tag{3.350}$$

Consider the second term on the left side of Eq. (3.349). It may be rewritten as

$$Xg_R\langle 0|\left(\partial_\mu\mathbf{c}_R^+\hat{}\mathbf{A}_{\mu R}\right)\delta\psi_R\bar{\psi}_R|0\rangle$$

$$= Xg_R\langle 0|\partial_\mu\left(\mathbf{c}_R^+\hat{}\mathbf{A}_{\mu R}\right)\delta\psi_R\bar{\psi}_R|0\rangle - Xg_R\langle 0|\mathbf{c}_R^+\hat{}(\partial_\mu\mathbf{A}_{\mu R})\delta\psi_R\bar{\psi}_R|0\rangle \tag{3.351}$$

$$= Xg_R\langle 0|\partial_\mu\left(\mathbf{c}_R^+\hat{}\mathbf{A}_{\mu R}\right)\delta\psi_R\bar{\psi}_R|0\rangle - \frac{1}{2}Xg_R\alpha_R\langle 0|\delta\left(\mathbf{c}_R^+\hat{}\mathbf{c}_R^+\right)\delta\psi_R\bar{\psi}_R|0\rangle \tag{3.352}$$

$$= Xg_R\langle 0|\partial_\mu\left(\mathbf{c}_R^+\hat{}\mathbf{A}_{\mu R}\right)\delta\psi_R\bar{\psi}_R|0\rangle + \frac{1}{2}Xg_R\alpha_R\langle 0|\delta\left(\mathbf{c}_R^+\hat{}\mathbf{c}_R^+\right)\delta\psi_R\bar{\psi}_R|0\rangle \tag{3.353}$$

In the last step we used both

$$\frac{d}{d\lambda_R}\langle 0|\left(\mathbf{c}_R^+\hat{}\mathbf{c}_R^+\right)\delta\psi_R\bar{\psi}_R|0\rangle = 0 \tag{3.354}$$

and the nilpotency

$$\delta^2\psi_R = 0 \tag{3.355}$$

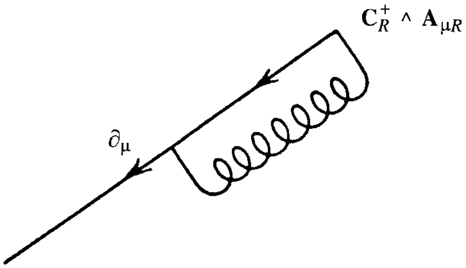



**Figure 3.12** Example of divergent graph for the operator $(\mathbf{c}_R^+ \hat{\ } \mathbf{A}_\mu)$.

To prove Eq. (3.355), proceed as follows:

$$\delta\psi = +g\psi\mathbf{T}\cdot\mathbf{c}\delta\lambda \tag{3.356}$$

$$\delta^2\psi = -gT\cdot\left(-\frac{g}{2}\mathbf{c}\hat{\ }\mathbf{c}\delta\lambda_2\right)\psi\delta\lambda_1 - g\mathbf{T}\cdot\mathbf{c}(-g\mathbf{T}\cdot\mathbf{c}\psi\delta\lambda_2)\delta\lambda_1 \tag{3.357}$$

$$= g^2\left(T_aT_b - \frac{1}{2}f_{cab}T^c\right)c^ac^b\psi\delta\lambda_2\delta\lambda_1 \tag{3.358}$$

$$= \frac{1}{2}g^2\left\{[T_a, T_b] - f_{abcT^c}\right\}c^ac^b\psi\delta\lambda_2\delta\lambda_1 = 0 \tag{3.359}$$

Substituting Eq. (3.353) into Eq. (3.350) and dividing by $X$ gives

$$\left\langle 0\left|\left[-\frac{\mathscr{Z}}{X}\,\square\,\mathbf{c}_R^+ + g_R\partial_\mu\left(\mathbf{c}_R^+\hat{\ }\mathbf{A}_{\mu R}\right)\right]\delta\psi_R\bar{\psi}_R\right|0\right\rangle$$
$$+\frac{g_R\alpha_R}{2}\langle 0|\left(\mathbf{c}_R^+\hat{\ }\mathbf{c}_R^+\right)\delta\psi_R\delta\bar{\psi}_R|0\rangle$$
$$= -ig_R\mathbf{T}\cdot\langle 0|\psi_R\bar{\psi}_R|0\rangle \tag{3.360}$$

Now $(\mathbf{c}_R^+\hat{\ }\mathbf{A}_{\mu R})$ is not a finite operator; an example of a divergent low-order graph is given in Fig. 3.12. But its counterterm is given by $(1-\mathscr{Z}/X)\partial_\mu c_R^+/g_R$. This can be seen by using the fact that $\delta\mathbf{c}_R$ and $\delta\mathbf{c}_R$ are finite, since by steps similar to the above we can obtain from Eq. (3.348),

$$ig_R\langle 0|\mathbf{T}\hat{\ }\mathbf{c}_R\mathbf{c}_R^+|0\rangle$$
$$= \left\langle 0\left|\left[g_R\partial_\mu\left(\mathbf{c}_R^+\hat{\ }\mathbf{A}_{\mu R}\right) - \frac{\mathscr{Z}}{X}\,\square\,\mathbf{c}_R^+ + \frac{\alpha_R g_R}{2}\mathbf{c}_R^+\hat{\ }\mathbf{c}_R^+\right]\Delta\mathbf{c}_R\mathbf{c}_R^+\right|0\right\rangle \tag{3.361}$$

$$= \langle 0| - \square c_R^+\delta c_R c_R^+|0\rangle$$
$$+\left\langle 0\left|\left[g_R\partial_\mu\left(\mathbf{c}_R^+\hat{\ }\mathbf{A}_{\mu R}\right) + \left(1-\frac{\mathscr{Z}}{X}\right)\square\,c_R^+\right]\delta c_R c_R^+\right|0\right\rangle$$
$$+\left\langle 0\left|\frac{\alpha_R g_R}{2}c_R^+\hat{\ }c_R^+\delta c_R\delta c_R^+\right|0\right\rangle \tag{3.362}$$

Using the form of this counterterm, we conclude from Eq. (3.360) that $\delta\psi_R$ is itself finite, as required.

There remains the vertex (Fig. 3.11*b*). For it, consider the matrix element

$$\langle 0|\psi_R \partial_\mu \mathbf{A}_{\mu R} \bar{\psi}_R|0\rangle = \alpha_R \langle 0|\psi_R \delta \mathbf{c}_R^+ \bar{\psi}|0\rangle \tag{3.363}$$

$$= \alpha_R \left[ -\langle 0|\delta\psi_R \mathbf{c}_R^+ \bar{\psi}|0\rangle - \langle 0|\psi_R \mathbf{c}_R^+ \delta\bar{\psi}|0\rangle \right] \tag{3.364}$$

which is finite, as required. This completes the demonstration that quantum chromodynamics is renormalizable.

For electroweak forces we must include scalar fields and handle the massive vectors created through the Higgs mechanism. The second problem is treated in Section 3.6. Let us here add scalars (to gauge vectors and fermions) with a coupling

$$\mathscr{L}_{\text{scalar}} = \frac{1}{2} D_\mu \phi D_\mu \phi - \frac{1}{2} M_s^2 \phi^2 - \frac{1}{3} f \phi^3 - \frac{1}{4} \lambda \phi^4 + g \bar{\psi} \psi \phi \tag{3.365}$$

where all labels on $\phi$ have been suppressed; $\phi$ may belong to an arbitrary representation of the gauge group, and its BRST transformation is

$$\delta\phi = g\mathbf{T} \cdot \mathbf{c}\phi\delta\lambda \tag{3.366}$$

We define renormalization constants $Z_s$, $X_\lambda$, $X_f$, $X_G$, and $Z_M^s$ according to

$$\phi = \sqrt{Z_s}\phi_R \tag{3.367}$$

$$\lambda = \frac{X_\lambda}{Z_s^2}\lambda_R \tag{3.368}$$

$$f = \frac{X_f}{\sqrt{Z_s}} f_R \tag{3.369}$$

$$G = \frac{X_G}{Z_2\sqrt{Z_s}} G_R \tag{3.370}$$

$$M_s = \sqrt{Z_m^s} M_{sR} \tag{3.371}$$

These five constants are defined such that the four Green's functions denoted in Fig. 3.13 are finite. Note that both $Z_s$ and $Z_m^s$ are used for Fig. 3.13*a*.

There remain the four possible Green's functions of Fig. 3.14 to be proved finite. The scalar–vector coupling (Fig. 3.14*d*) occurs only if the scalars are in the adjoint representation. Note that the three cases shown in Fig. 3.15, although they have a superficial degree of divergence $D = 0$, are rendered finite by the external derivative in the case of Fig. 3.15*a* and by Lorentz invariance (requiring an external $p_\mu$) in the cases 3.15*b* and *c*.

When scalars are present, we must return to the proof of renormalizability for pure Yang–Mills theory. In step 1, particularly Eqs. (3.330) and (3.331), there are

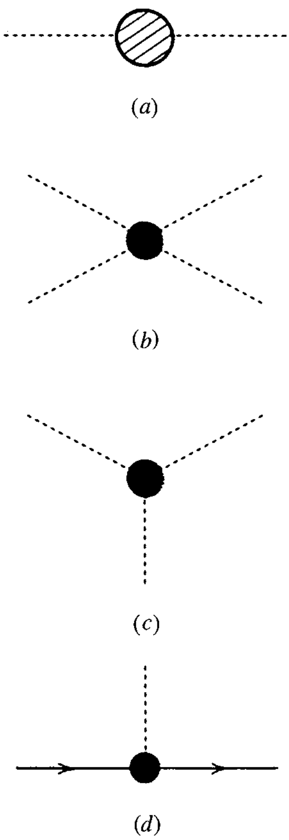



**Figure 3.13** Four Green's functions made finite by definitions of $Z_s$, $X_\lambda$, $X_f$, $X_G$, and $Z_m^s$.

now three combinations of fields: $\mathbf{c}_R^+$, $\mathbf{c}_R^+\mathbf{A}_{\mu R}$, and $\mathbf{c}_R^+\phi$, which can make the left-hand side of Eq. (3.327) divergent, but the proof remains, and steps 2 through 5 are also unchanged.

Next we look at $d\phi_R/d\lambda_R$, given by

$$d\phi_R = (Xg_R\phi_R\mathbf{T}\cdot\mathbf{c}_R)d\lambda_R \tag{3.372}$$

That this operator is divergent can be seen by Fig. 3.16, for example. The procedure is similar to that for the fermion case; namely, one uses the ghost equation of motion and considers the relations

$$\left\langle 0\left|\frac{\delta\mathscr{L}}{\delta\mathbf{c}_R}\delta\phi_R\right|0\right\rangle = i\left\langle 0\left|\frac{\delta}{\delta\mathbf{c}_R}(\delta\phi_R)\right|0\right\rangle \tag{3.373}$$

$$\left\langle 0\left|\frac{\delta\mathscr{L}}{\delta\mathbf{c}_R}\delta\phi_R\phi_R\right|0\right\rangle = i\left\langle 0\left|\frac{\delta}{\delta\mathbf{c}_R}(\delta\phi_R\phi_R)\right|0\right\rangle \tag{3.374}$$

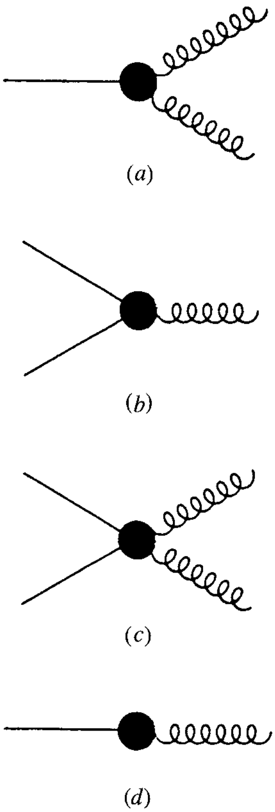



Figure 3.14 Four extra divergences when scalars are added to gauge theory with fermions.

Using the counterterm derived previously in the fermion proof, one finds that

$$g_R\left\langle 0\left|\left[\partial_\mu(\mathbf{A}_{\mu R}\,\hat{}\,\mathbf{c}_R)+\frac{1}{g_R}\left(\frac{\mathscr{Z}}{X}-1\right)\Box\,\mathbf{c}_R^{+}\right]\delta\phi_R\right|0\right\rangle \tag{3.375}$$

and

$$g_R\left\langle 0\left|\left[\partial_\mu(\mathbf{A}_{\mu R}\,\hat{}\,\mathbf{c}_R)+\frac{1}{g_R}\left(\frac{\mathscr{Z}}{X}-1\right)\Box\,\mathbf{c}_R^{+}\right]\delta\phi_R\phi_R\right|0\right\rangle \tag{3.376}$$

are finite, and hence $\delta\phi_R$ itself.

To show finiteness of the four Green's functions in Fig. 3.14 is a simple exercise. We consider the four BRST identities

$$\frac{d}{d\lambda_R}\langle 0|T\left(c_R^{+}\phi_R A_{\mu R}\right)|0\rangle=0 \tag{3.377}$$

$$\frac{d}{d\lambda_R}\langle 0|T\left(\phi_R\phi_R c_R^{+}\right)|0\rangle=0 \tag{3.378}$$

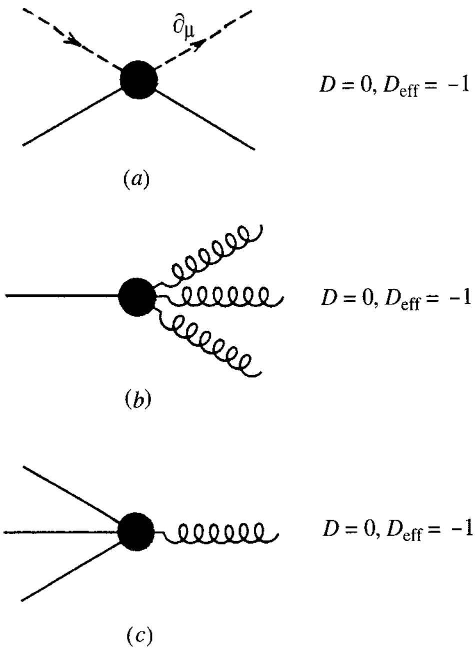



**Figure 3.15** Three vertices rendered finite by external derivative or Lorentz invariance.

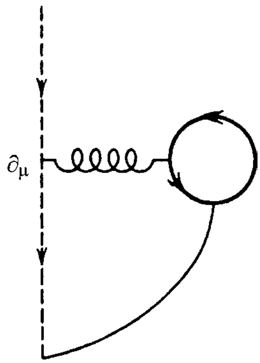



**Figure 3.16** Example of divergent diagram for $d\phi_R/d\lambda_R$.

$$\frac{d}{d\lambda_R}\langle 0|T\left(\phi_R\phi_R c_R^+ A_{\mu R}\right)|0\rangle = 0 \tag{3.379}$$

$$\frac{d}{d\lambda_R}\langle 0|T\left(c_R^+\phi_R\right)|0\rangle = 0 \tag{3.380}$$

respectively, and the required result follows. This completes the proof for electroweak dynamics, still with massless vectors, except for one loop hole, as follows.

Since we have been tacitly assuming the use of dimensional regularization everywhere, closed fermion loops that involve $\gamma_5$ have been excluded. In a general Feynman diagram, the closed fermion loops are necessarily disjoint and can apparently be treated one at a time. Bardeen [50] suggested the following procedure:

1. Dimensionally regularize all boson integrations.
2. Perform the Dirac traces in four dimensions.
3. Regularize the fermion loop momentum integrals using either dimensional or Pauli–Villars techniques.

This prescription is suggested as preserving all Ward identities, provided that the anomalies are canceled, as discussed in Section 3.3. No detailed proof of the Bardeen prescription has been provided; this is a logical gap in the proof for renormalization of a chiral gauge theory [104].

## 3.6
## 'T Hooft Gauges

In the discussion so far, we have assumed that the vector propagator has the renormalizable behavior $\sim 1/p^2$, $p^2 \to \infty$. With massive vectors, this is not true, in general, and power counting is lost because divergences can grow arbitrarily by adding new internal massive vector lines to any given diagram.

As pointed out by 't Hooft [67, 68], the combination of non-Abelian local gauge invariance and spontaneous symmetry breaking (i.e., Higgs–Kibble mechanism) gets out of this impasse. Let us progress by stages, to obtain a clear picture of how this happens.

In an Abelian gauge theory, with only neutral vectors coupled to a conserved current, the addition of a mass term is harmless. For supposing that we take massive quantum electrodynamics with Lagrangian

$$\mathscr{L} = -\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 + \frac{1}{2}M^2 A_\mu A_\mu + \bar{\psi}(i\not{D} - M)\psi \tag{3.381}$$

The vector propagator is

$$-i\frac{g_{\mu\nu} - k_\mu k_\nu/M^2}{k^2 - M^2 + i\epsilon} \tag{3.382}$$

But if we decompose the vector field $A_\mu$ into transverse and longitudinal parts,

$$A_\mu(x) = a_\mu(x) + \frac{1}{M}\partial_\mu b(x) \tag{3.383}$$

The pairings give

$$a_\mu a_\nu \sim g_{\mu\nu} \tag{3.384}$$

$$\partial_\mu b \partial_\nu b \sim k_\mu k_\nu \tag{3.385}$$

in the propagator. Recall that the $S$-matrix depends not on $\mathscr{L}(x)$ but on $S = \int d^4x \mathscr{L}(x)$. Since $\mathscr{L}(x)$ contains $\bar{\psi}\gamma_\mu\psi A_\mu$, the action contains

$$\int d^4x \left( \bar{\psi}\gamma_\mu\psi a_\mu + \frac{1}{M}\bar{\psi}\gamma_\mu\psi\partial_\mu b \right) \tag{3.386}$$

Integrating the second term by parts and using the current conservation

$$\partial_\mu(\bar{\psi}\gamma_\mu\psi) = 0 \tag{3.387}$$

we see that the $k_\mu k_\nu/M^2$ part of Eq. (3.382) always gives zero contribution, and the theory is renormalizable.

In a non-Abelian theory the self-couplings of $\mathbf{A}_\mu$ give a new (gauge-variant) current, coupling to the vector field, which is not, in general, conserved and the argument above fails.

This simplest place to study 't Hooft gauge is the Abelian Higgs model [cf. Eqs. (1.274) through (1.282)]. The Lagrangian is

$$\mathscr{L} = \left|(\partial_\mu\phi + ieA_\mu\phi)\right|^2 - M_s^2\phi^*\phi - \frac{\lambda}{6}(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F_{\mu\nu} \tag{3.388}$$

Parametrizing complex $\phi$ as the two real fields $\zeta$ and $\eta$ gives

$$\phi = e^{i\zeta/v}\frac{v+\eta}{\sqrt{2}} \tag{3.389}$$

$$v = \sqrt{-\frac{6M_s^2}{\lambda}} \tag{3.390}$$

one finds (recall that $M^2 < 0$)

$$\begin{aligned} \mathscr{L} &= -\frac{1}{4}F_{\mu\nu}F_{\mu\nu} + \frac{1}{2}\partial_\mu\eta\partial_\mu\eta + M_s^2\eta^2 \\ &\quad + \frac{1}{2}e^2v^2A_\mu A_\mu + \frac{1}{2}\partial_\mu\zeta\partial_\mu\zeta + evA_\mu\partial_\mu\zeta \\ &\quad + \text{higher orders} \end{aligned} \tag{3.391}$$

Now choose the gauge-fixing term

$$\mathscr{L}_{\text{GF}} = -\frac{1}{2\alpha}(\partial_\mu A_\mu - \alpha ev\zeta)^2 \tag{3.392}$$

As for the covariant gauges considered in Chapter 2, this may be incorporated into the functional integral by writing the functional delta function to include the auxiliary function $f(x)$:

$$\int Df \prod_x \delta(\partial_\mu A_\mu - \alpha ev\zeta - f)\exp\left(-\frac{1}{2\alpha}\int d^4x f^2\right)\cdots \tag{3.393}$$

There is also the ghost term in $\mathscr{L}_{\text{eff}}$, of course, but that is irrelevant to the considerations of the present section. One finds ($M = ev$) that

$$\mathscr{L} + \mathscr{L}_{\text{GF}} = -\frac{1}{4}F_{\mu\nu}F_{\mu\nu} + \frac{1}{2}\partial_\mu\eta\partial_\mu\eta + M_s^2\eta^2$$
$$+ \frac{1}{2}M^2 A_\mu A_\mu + \frac{1}{2}\partial_\mu\zeta\partial_\mu\zeta - \frac{1}{2}\alpha M^2\zeta^2$$
$$- \frac{1}{2\alpha}(\partial_\mu A_\mu)^2 + \text{higher orders} \tag{3.394}$$

The terms quadratic in $A_\mu$ are given by

$$\frac{1}{2}A_\mu A_\nu M^{\mu\nu} \tag{3.395}$$

$$M^{\mu\nu} = M^2 g_{\mu\nu} - \frac{1}{\alpha}k_\mu k_\nu - k^2 g_{\mu\nu} + k_\mu k_\nu \tag{3.396}$$

$$= -(k^2 - M^2)\left(g_{\mu\nu} - \frac{k_\mu k_\nu}{k^2}\right)$$
$$+ \left(M^2 - \frac{1}{\alpha}k^2\right)\frac{k_\mu k_\nu}{k^2} \tag{3.397}$$

The inverse of $M_{\mu\nu}$ is given by $N_{\mu\nu}$ such that

$$M_{\mu\alpha}N_{\alpha\nu} = g_{\mu\nu} \tag{3.398}$$

One finds easily that

$$N_{\mu\nu} = -\frac{1}{k^2 - M^2}\left[g_{\mu\nu} - (1-\alpha)\frac{k_\mu k_\nu}{k^2 - \alpha M^2}\right] \tag{3.399}$$

The vector propagator is thus $iN_{\mu\nu}$. The crucial point is that for any finite $\alpha < \infty$, the propagator behaves $\sim 1/k^2$, $k^2 \to \infty$, just as in the massless case.

We also have the scalar $\eta$ with propagator

$$\frac{i}{k^2 + 2M^2 + i\epsilon} \tag{3.400}$$

and $\zeta$ with propagator

$$\frac{i}{k^2 - \alpha M^2} \tag{3.401}$$

Thus we have five states $A_\mu$ and $\zeta$ to discuss three components of $A_\mu$. As $\alpha \to \infty$, these are reduced to three because we have a genuine spin-1 $A_\mu$ and $M \to \infty$. The fields $A_\mu$ and $\zeta$ separately have no significance in the 't Hooft gauge. The poles at $k^2 = \alpha M^2$ in the propagators (3.399) and (3.401) can be shown to cancel in the $S$-matrix such that unitarity is preserved, although we do not present the proof here.

From Eq. (3.399) we observe that $\alpha \to 0$ corresponds to the Landau gauge; $\alpha \to 1$, to the Feynman gauge; and $\alpha \to \infty$, to the unitary gauge (discussed in Chapter 1). In the unitary gauge, the particle content, and unitarity, are manifest, but power-counting renormalizability is lost, since the vector propagator $\sim$ constant, $k^2 \to \infty$.

Note that when we include the interaction terms, there are many additional vertices to be accommodated in the Feynman rules, since

$$\begin{aligned}\mathscr{L} + \mathscr{L}_{\text{GF}} = {} & \text{(quadratic terms)} \\ & + \frac{1}{v^2}\big[(\partial_\mu \zeta)(\partial_\mu \zeta + M A_\mu)\big]\eta(2v + \eta) \\ & - \frac{1}{6}\lambda_0 \eta^3 - \frac{1}{24}\lambda_0 \eta^4 + \frac{1}{2}e^2 A_\mu A_\mu \eta(2v + \eta) \end{aligned} \tag{3.402}$$

The vertices are $(+i) \times \mathscr{L}_{\text{int}}$ in the Feynman rules, plus appropriate combinatorial $(p!)$ factors for identical fields.

Next consider an $O(3)$ non-Abelian Higgs model [cf. Eqs. (1.286) through (1.292)]. The Lagrangian is

$$\mathscr{L} = \frac{1}{2}\big(\partial_\mu \phi_k + f \epsilon_{klm} A_\mu^l \phi^m\big)^2 - V(\phi_k \phi_k) - \frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a \tag{3.403}$$

In this we parametrize the triplet $\phi^a$ by

$$\phi = \exp\left[-\frac{i}{v}(\xi_1 T_1 + \xi_2 T_2)\right]\begin{pmatrix} 0 \\ 0 \\ v + \eta \end{pmatrix} \tag{3.404}$$

$$\simeq \begin{pmatrix} \xi_2 \\ -\xi_1 \\ v + \eta \end{pmatrix} + \text{higher orders} \tag{3.405}$$

where we used $(T^i)_{jk} = -i\epsilon_{ijk}$. Substitution into Eq. (3.403) gives, after some algebra,

$$\begin{aligned}\mathscr{L} = {} & \frac{1}{2}\big[(\partial_\mu \xi_1)^2 + (\partial_\mu \xi_2)^2 + (\partial_\mu \eta)^2\big] + M_s^2 \eta^2 \\ & + \left(\frac{1}{2}M^2 + g^2 v\eta + \frac{1}{2}g^2\eta^2\right)\big(A_\mu^1 A_\mu^1 + A_\mu^2 A_\mu^2\big) \\ & + (M + g\eta)\big(A_\mu^1 \partial_\mu \xi^1 + A_\mu^2 \partial_\mu \xi^2\big) - \frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a \end{aligned}$$

$$-\frac{1}{4}M_s^2 v^2 + \frac{3M_s^2}{2v}\eta^3 + \frac{M_s^2}{4v^2}\eta^4 + \frac{M_s^2}{v}\xi^2\eta$$

$$+\frac{M_s^2}{2v^2}\boldsymbol{\xi}^2\eta^2 + \frac{M_s^2}{2v^2}\boldsymbol{\xi}^4 \tag{3.406}$$

where $\boldsymbol{\xi}^2 = \xi_1^2 + \xi_2^2$. To go to the 't Hooft gauge we use as the gauge-fixing term,

$$\mathscr{L}_{\text{GF}} = -\frac{1}{2\alpha}\left(\partial_\mu A_\mu^a - \alpha M \xi^a\right)^2 \tag{3.407}$$

with $M = gv$. The $A_\mu^{1,2} - \xi^{1,2}$ mixing terms then cancel (after integration by parts) and the propagators are seen to be

$$A_\mu^3 : \frac{-ig_{\mu\nu}}{k^2 + i\epsilon} \tag{3.408}$$

$$A_\mu^{1,2} : \frac{-i[g_{\mu\nu} - (1-\alpha)k_\mu k_\nu/(k^2 - \alpha M^2)]}{k^2 - M^2 + i\epsilon} \tag{3.409}$$

$$\xi^{1,2} : \frac{+i}{k^2 - \alpha M^2} \tag{3.410}$$

$$\eta : \frac{i}{k^2 + 2M^2 + i\epsilon} \tag{3.411}$$

The propagators for the massive $A_\mu^{1,2}$ are seen to be renormalizable, behaving as $k^{-2}$, $|k| \to \infty$.

As a final, more realistic example, consider the Glashow–Weinberg–Salam model. We seek its gauge-fixing term. The part of the Lagrangian involving Higgs scalars is

$$\mathscr{L}_s = \left(\partial_\mu \phi^+ + \frac{ig'}{2}B_\mu \phi^+ + \frac{ig}{2}\tau^i A_\mu^i \phi^+\right)$$

$$\cdot\left(\partial_\mu \phi - \frac{ig'}{2}B_\mu \phi - \frac{ig}{2}\tau^i A_\mu^i \phi\right) - \mu^2 \phi^+ \phi - \lambda\left(\phi^+\phi\right)^2 \tag{3.412}$$

Parametrize

$$\phi = \left(1 - \frac{i}{v}\boldsymbol{\tau}\cdot\boldsymbol{\xi}\right)\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + \eta \end{pmatrix} \tag{3.413}$$

$$= \begin{pmatrix} -\frac{i}{v}\xi_1 - \frac{i}{v}\xi_2 \\ 1 - \frac{i}{v}\xi_3 \end{pmatrix}\frac{1}{\sqrt{2}}(v + \eta) \tag{3.414}$$

Then we find that $\mathscr{L}$ contains ($M_B = \frac{1}{2}g'v$, $M_W = \frac{1}{2}gv$)

$$-M_B B_\mu \partial_\mu \xi_3 + M_W \mathbf{A}_\mu \cdot \partial_\mu \boldsymbol{\xi} + \frac{1}{2}\partial_\mu \boldsymbol{\xi} \cdot \partial_\mu \boldsymbol{\xi} \tag{3.415}$$

Thus we define the gauge fixing as

$$\mathscr{L}_{\text{GF}} = -\frac{1}{2\alpha}\left(\partial_\mu A^a_\mu - \alpha M_W \xi^a\right)^a - \frac{1}{2a}\left(\partial_\mu B_\mu + \alpha M_B \xi^3\right)^2 \tag{3.416}$$

The masses are now given by the term

$$-\frac{1}{2}\alpha M_W^2\left(\xi_1^2 + \xi_2^2\right) - \frac{1}{2}\alpha M_Z^2 (\xi_3)^2 \tag{3.417}$$

where we used

$$M_W^2 + M_B^2 = M_Z^2 \tag{3.418}$$

Rewriting in terms of physical fields by

$$A^3_\mu = -\sin\theta_W A_\mu + \cos\theta_W Z_\mu \tag{3.419}$$

$$B_\mu = \cos\theta_W A_\mu + \sin\theta_W Z_\mu \tag{3.420}$$

then

$$\begin{aligned}\mathscr{L}_{\text{GF}} &= -\frac{1}{2\alpha}\sum_{i=1}^{2}\left(\partial_\mu A^i_\mu - \alpha M_W \xi^i\right)^2 - \frac{1}{2\alpha}(\partial_\mu A_\mu)^2 \\ &\quad - \frac{1}{2\alpha}\left(\partial_\mu z_\mu + \alpha M_Z \xi^3\right)^2\end{aligned} \tag{3.421}$$

Combining this with the quadratic terms

$$\begin{aligned}&-\frac{1}{4}\left(\partial_\mu A^i_\nu - \partial_\nu A^i_\mu\right)^2 - \frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2 \\ &\quad + \frac{1}{2}M_W^2\left(A^1_\mu A^1_\mu A^2_\mu A^2_\mu\right) + \frac{1}{2}M_Z^2 Z_\mu Z_\mu\end{aligned} \tag{3.422}$$

one deduces the set of propagators

$$A_\mu : \frac{-i g_\mu \nu}{k^2 + i\epsilon} \tag{3.423}$$

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(A^1_\mu \pm i A^2_\mu\right) : \frac{-i[g_{\mu\nu} - (1-\alpha)k_\mu k_\nu/(k^2 - \alpha M_W^2)]}{k^2 - M_W^2 + i\epsilon} \tag{3.424}$$

$$Z_\mu : \frac{-i[g_{\mu\nu} - (1-\alpha)k_\mu k_\nu/(k^2 - \alpha M_Z^2)]}{k^2 - M_Z^2 + i\epsilon} \tag{3.425}$$

$$\xi^{1,2} : \frac{i}{k^2 - \alpha M_W^2} \tag{3.426}$$

$$\xi^3 : \frac{i}{k^2 - \alpha M_Z^2} \tag{3.427}$$

To establish unitarity, it is necessary to demonstrate that the unphysical ($\alpha$-dependent) poles cancel. This is shown in detail in Refs. [74] and [86] (note that the $R_\xi$ gauges therein coincide with the above, with $\xi = 1/\alpha$).

Using such gauges, the renormalizability of the non-Abelian Higgs models is demonstrated. But at this point the reader may well raise the question of uniqueness. For the models we have considered, when written in the U-gauge, renormalizability is not obvious and it is only after special manipulations that it becomes so. Thus how do we know that the Higgs models are not just special cases of apparently nonrenormalizable models that can eventually be rewritten, with sufficient cleverness, in a manifestly renormalizable form?

There is a plausibility argument for uniqueness based on tree unitarity [105–112]. Consider the $N$-point tree amplitude, and suppose that all momentum transfers and subenergies grow $\sim E^2$, $E \to \infty$ with nonexceptional momenta (i.e., all Euclidean), or at least away from Landau singularities. Then tree unitarity requires that

$$T_{N-n,n} < E^{4-N}, \qquad E \to \infty \tag{3.428}$$

To see the reason for this, consider the unitary relation

$$\int d\Omega_{N-2} |T_{N-2,2}|^2 < \mathrm{Im}\, T_{2,2} < \text{constant} \tag{3.429}$$

(neglecting logarithms). Since the phase space $\Omega_{N-2} \sim E^{2N-8}$, we find Eq. (3.428) for consistency. The precise connection between unitarity and renormalization is unknown, so it is only a conjecture that tree unitarity is necessary.

It can be shown that Higgs models preserve tree unitarity [112]. The inverse problem can also be solved: What constraints are imposed on the couplings of spin 0, $\frac{1}{2}$, and 1 by tree unitarity? The answer is that there must be Yang–Mills couplings, and for non-Abelian massive vectors, there must be Higgs scalars coupling precisely according to the gauge theory model.

The algebraic detail is provided in the original papers cited [105–112] and is not very illuminating in itself. Two examples may be mentioned.

**Example 1.** $\nu + \bar{\nu} \to W^+ + W^-$ **(Longitudinal Helicities)**. In this case, the electron exchange diagram (Fig. 3.17*a*) violates tree unitarity, but when combined with the neutral current boson diagram (Fig. 3.17*b*), the amplitude behaves correctly.

**Example 2.** $e^+ + e^- \to W^+ + W^-$. Here, the anomalous magnetic moment of $W$ makes the electromagnetic pair production badly behaved (Fig. 3.18*a*). All four diagrams add to restore tree unitarity.

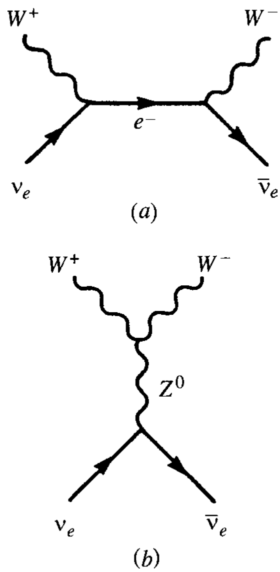



**Figure 3.17** Two Born diagrams for $\nu\bar{\nu} \to W^+W^-$.

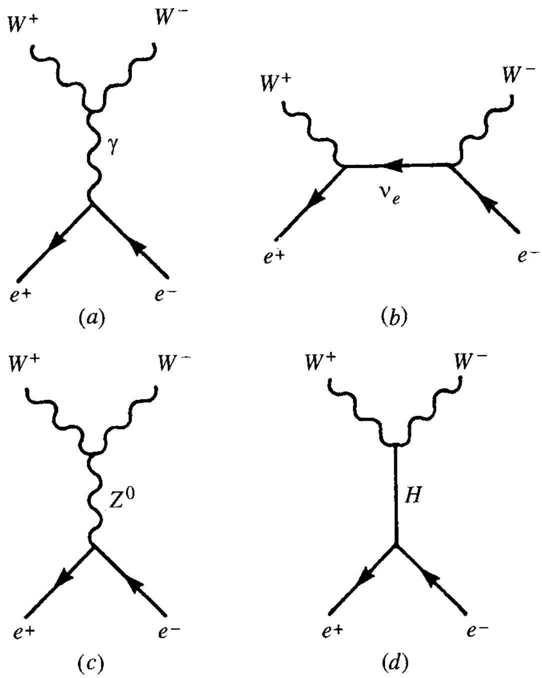



**Figure 3.18** Four Born diagrams for $e^+e^- \to W^+W^-$.

The existence of this argument makes it likely that the only renormalizable models of charged and massive vectors are gauge theories.

Note that no field theories with a finite number of spins greater than 1 are known to be renormalizable. For example, all extended supergravity theories are believed to have uncontrollable divergences at the three-loop level.

## 3.7
## Summary

To arrive at our present stage, we have used four key ideas, which we may state briefly:

1. The combination of local gauge invariance and spontaneous breaking in the Higgs mechanism. This introduces scalar fields.
2. The introduction of fictitious ghost fields (Faddeev–Popov) in quantization and the Feynman rules.
3. The use of ghost fields in the gauge function to simplify proof of the Ward identities and renormalizability (BRST).
4. The choice of gauge relating divergence of the gauge field to the Higgs scalar field, facilitating proof of renormalization after spontaneous breaking ('t Hooft).

The payoff from all this mathematics is a new wide class of renormalizable theories, on a footing with QED, and one must attempt to use them for describing the fundamental forces. For strong interactions, the standard idea is to use exact Yang–Mills with an octet of SU(3) gauge gluons coupling to color triplet quarks. We have shown that this theory is ultraviolet renormalizable; the further important ultraviolet property of asymptotic freedom is discussed in Chapters 5 and 6.

For unified weak and electromagnetic interactions, one chooses the gauge group to be $SU(2) \times U(1)$, with appropriate representations for the fermion (lepton and quark) matter fields. The Higgs scalars are put in doublets of the SU(2), and one can then successfully calculate higher-order weak interactions, as described in Chapter 4.

### References


1. W. Pauli and F. Villars, *Rev. Mod. Phys.* 21, 433 (1949).
2. J. Rayski, *Acta Phys. Pol.* 9, 129 (1948).
3. J. Rayski, *Phys. Rev.* 75, 1961 (1949).
4. E.C.G. Stückelberg and D. Rivier, *Phys. Rev.* 74, 218 (1948).
5. E.C.G. Stückelberg and D. Rivier, *Phys. Rev.* 74, 986 (1948).
6. C.G. Bollini and J.J. Giambiagi, *Phys. Lett.* 40B, 566 (1972).
7. C.G. Bollini and J.J. Giambiagi, *Nuovo Cimento* 31, 550 (1964).
8. G.M. Cicuta and E. Montaldi, *Lett. Nuovo Cimento* 4, 392 (1972).
9. P. Butera, G.M. Cicuta and E. Montaldi, *Nuovo Cimento* 19A, 513 (1974).
10. J.F. Ashmore, *Lett. Nuovo Cimento* 4, 289 (1972).
11. G. 't Hooft and M. Veltman, *Nucl. Phys.* B44, 189 (1972).

12 G. 't Hooft, *Nucl. Phys.* B61, 455 (1973).
13 G. 't Hooft, *Nucl. Phys.* B62, 444 (1973).
14 G. 't Hooft and M. Veltman, CERN Report 73-9 Diagrammar, 1973.
15 E.R. Speer, *J. Math. Phys.* 9, 1404 (1968).
16 E.R. Speer, *Commun. Math. Phys.* 23, 23 (1971).
17 R.P. Feynman, *Phys. Rev.* 76, 769 (1949).
18 A. Erdelyi et al., *Higher Transcendental Functions: Bateman Manuscript Project*, McGraw–Hill, New York, 1953.
19 J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics*, McGraw–Hill, New York, 1964.
20 E.A. Uehling, *Phys. Rev.* 48, 55 (1935).
21 W.E. Lamb and R.C. Retherford, *Phys. Rev.* 72, 241 (1947).
22 B.E. Lautrup, A. Peterman and E. de Rafael, *Phys. Rev.* 3C, 193 (1972).
23 J.C. Ward, *Phys. Rev.* 78, 182 (1950).
24 J.C. Ward, *Proc. Phys. Soc.* 64, 54 (1951).
25 Y. Takahashi, *Nuovo Cimento* 6, 371 (1957).
26 J. Schwinger, *Phys. Rev.* 73, 416 (1948).
27 P. Kusch and H. Foley, *Phys. Rev.* 72, 1256 (1947).
28 P. Kusch and H. Foley, *Phys. Rev.* 73, 412 (1948).
29 R.S. van Dykk, P.B. Schwinberg and H.G. Dehmelt, *Phys. Rev. Lett.* 59, 26 (1987).
30 T. Aoyama, M. Hayakawa, T. Kinoshita and M. Nio, arXiv:0712.2607 [hep-ph].
31 D.A. Akyeampong and R. Delbourgo, *Nuovo Cimento* 17A, 578 (1973).
32 D.A. Akyeampong and R. Delbourgo, *Nuovo Cimento* 18A, 94 (1973).
33 D.A. Akyeampong and R. Delbourgo, *Nuovo Cimento* 19A, 219 (1974).
34 P. Breitenlohner and D. Maison, *Commun. Math. Phys.* 52, 11 (1977).
35 P. Breitenlohner and D. Maison, *Commun. Math. Phys.* 52, 39 (1977).
36 P. Breitenlohner and D. Maison, *Commun. Math. Phys.* 52, 55 (1977).
37 C. Becchi, A. Rouet and R. Stora, *Commun. Math. Phys.* 42, 127 (1975). L.V. Tyutin, Lebedev Inst. preprint no. 39 (1975).
38 J.C. Taylor, *Nucl. Phys.* B33, 436 (1971).
39 A.A. Slavnov, *Theor. Math. Phys.* 10, 99 (1972).
40 J. Steinberger, *Phys. Rev.* 76, 1180 (1949).
41 J. Schwinger, *Phys. Rev.* 82, 664 (1951).
42 S.L. Adler, *Phys. Rev.* 177, 2426 (1969).
43 J.S. Bell and R. Jackiw, *Nuovo Cimento* 60A, 47 (1969).
44 D.G. Sutherland, *Phys. Lett.* 23, 384 (1966).
45 D.G. Sutherland, *Nucl. Phys.* B2, 433 (1967).
46 C. Bouchiat, J. Iliopoulos and P. Meyer, *Phys. Lett.* 38B, 519 (1972).
47 D.J. Gross and R. Jackiw, *Phys. Rev.* D6, 477 (1972).
48 H. Georgi and S.L. Glashow, *Phys. Rev.* D6, 429 (1972).
49 W.A. Bardeen, *Phys. Rev.* 184, 1848 (1969).
50 W.A. Bardeen, *Proceedings of the 16th International Conference on High Energy Physics*, Fermilab, 1972, Vol. 2, p. 295.
51 W.A. Bardeen, *Nucl. Phys.* B75, 246 (1974).
52 S.L. Adler and W.A. Bardeen, *Phys. Rev.* 182, 1517 (1969).
53 K.G. Wilson, *Phys. Rev.* 181, 1909 (1969).
54 R. Aviv and A. Zee, *Phys. Rev.* D5, 2372 (1972).
55 S.L. Adler, C.G. Callan, D.J. Gross and R. Jackiw, *Phys. Rev.* D6, 2982 (1972).
56 J. Wess and B. Zumino, *Phys. Rev.* 37B, 95 (1971).
57 R.W. Brown, C.-G. Shih and B.-L. Young, *Phys. Rev.* 186, 1491 (1969).
58 S.L. Adler, in *Lectures on Elementary Particles and Quantum Field Theory*, Brandeis University, S. Deser, M. Grisaru, and H. Pendleton, eds., vol. 1, MIT Press, Cambridge, Mass., 1970, p. 1.
59 R. Jackiw, in *Lectures on Current Algebra and Its Applications*, S.B. Treiman, R. Jackiw, and D.J. Gross, eds., Princeton University Press, Princeton, NJ, 1972, p. 94.
60 M. Veltman, *Proc. R. Soc.* A301, 107 (1967).
61 Particle Data Group, *Rev. Mod. Phys.* 56, S1 (1984).
62 G. Curci and R. Ferrari, *Nuovo Cimento* 32A, 151 (1976).
63 G. Curci and R. Ferrari, *Phys. Lett.* 63B, 91 (1976).
64 I. Ojima, *Prog. Theor. Phys.* 64, 625 (1980).
65 L. Baulieu and J. Thierry-Mieg, *Nucl. Phys.* B197, 477 (1982).

66 L. Alvarez-Gaume and L. Baulieu, *Nucl. Phys.* B212, 255 (1983).
67 G. 't Hooft, *Nucl. Phys.* B33, 173 (1971).
68 G. 't Hooft, *Nucl. Phys.* B35, 167 (1971).
69 G. 't Hooft and M. Veltman, *Nucl. Phys.* B50, 318 (1972).
70 B.W. Lee, *Phys. Rev.* D5, 823 (1972).
71 B.W. Lee and J. Zinn-Justin, *Phys. Rev.* D5, 3121 (1972).
72 B.W. Lee and J. Zinn-Justin, *Phys. Rev.* D5, 3137 (1972).
73 B.W. Lee and J. Zinn-Justin, *Phys. Rev.* D5, 3155 (1972).
74 B.W. Lee and J. Zinn-Justin, *Phys. Rev.* D7, 1049 (1978).
75 B.W. Lee, *Phys. Lett.* 46B, 214 (1973).
76 B.W. Lee, *Phys. Rev.* D9, 933 (1974).
77 S.D. Joglekar and B.W. Lee, *Ann. Phys.* 97, 160 (1976).
78 D.A. Ross and J.C. Taylor, *Nucl. Phys.* B51, 125 (1973).
79 H. Kluberg-Stern and J.B. Zuber, *Phys. Rev.* D12, 467 (1975).
80 H. Kluberg-Stern and J.B. Zuber, *Phys. Rev.* D12, 482 (1975).
81 H. Kluberg-Stern and J.B. Zuber, *Phys. Rev.* D12, 3159 (1975).
82 T. Appelquist and H. Quinn, *Phys. Lett.* 39B, 229 (1972).
83 W.A. Bardeen, R. Gastmans and B. Lautrup, *Nucl. Phys.* B46, 319 (1972).
84 T.W. Appelquist, J.R. Primack and H.R. Quinn, *Phys. Rev.* D6, 2998 (1972).
85 J.R. Primack and H.R. Quinn, *Phys. Rev.* D6, 2923 (1972).
86 K. Fujikawa, B.W. Lee and A.I. Sanda, *Phys. Rev.* D6, 2923 (1972).
87 T.W. Appelquist, J.R. Primack and H.R. Quinn, *Phys. Rev.* D7, 2998 (1973).
88 T.W. Appelquist, J. Carazzone, T. Goldman and H.R. Quinn, *Phys. Rev.* D8, 1747 (1973).
89 B.W. Lee, *Nucl. Phys.* B9, 649 (1969).
90 J.L. Gervais and B.W. Lee, *Nucl. Phys.* B12, 627 (1969).
91 B.W. Lee, *Chiral Dynamics*, Gordon and Breach, New York, 1972.
92 K. Symanzik, *Nuovo Cimento Lett.* 1, 10 (1969).
93 K. Symanzik, *Commun. Math. Phys.* 16, 48 (1970).
94 E.S. Abers and B.W. Lee, *Phys. Rep.* 9C, 1 (1973).
95 S. Coleman, in *Laws of Hadronic Matter*, A. Zichichi, ed., Academic Press, New York, 1975, p. 139.
96 G. Costa and M. Tonin, *Rev. Nuovo Cimento* 5, 29 (1975).
97 A.A. Slavnov, *Sov. J. Part. Nucl.* 5, 303 (1975).
98 B.W. Lee, in *Methods in Field Theory*, R. Balian and J. Zinn-Justin, eds., North-Holland, Amsterdam, 1976, p. 79.
99 R.P. Feynman, in *Weak and Electromagnetic Interactions at High Energy*, R. Balian and C.H. Llewellyn Smith, eds., North-Holland, Amsterdam 1977, p. 121.
100 J.C. Taylor, *Gauge Theories of Weak Interactions*, Cambridge University Press, Cambridge, 1976.
101 S. Weinberg, *Phys. Rev. Lett.* 19, 1264 (1967).
102 A. Salam, in *Elementary Particle Theory*, N. Svartholm, ed., Almqvist & Wiksell, Stockholm, 1968, p. 367.
103 J.C. Collins, *Renormalization*, Cambridge University Press, Cambridge, 1984.
104 P.H. Frampton, *Phys. Rev.* D20, 3372 (1979).
105 S. Weinberg, *Phys. Rev. Lett.* 27, 1688 (1971).
106 J. Schechter and Y. Ueda, *Phys. Rev.* D7, 3119 (1973).
107 J.M. Cornwall, D.N. Levin and G. Tiktopoulos, *Phys. Rev. Lett.* 30, 1268 (1973).
108 C.H. Llewellyn Smith, *Phys. Lett.* 46B, 233 (1973).
109 J.M. Cornwall, D.N. Levin and G. Tiktopoulos, *Phys. Rev. Lett.* 32, 498 (1974).
110 J.M. Cornwall, D.N. Levin and G. Tiktopoulos, *Phys. Rev.* D10, 1145 (1974).
111 D.N. Levin and G. Tiktopoulos, *Phys. Rev.* D12, 415 (1975).
112 J.S. Bell, *Nucl. Phys.* B60, 427 (1973).

# 4
# Electroweak Forces

## 4.1
## Introduction

As an example of a spontaneously broken gauge theory, we discuss the standard SU(2) × U(1) electroweak theory. This theory contains many free parameters, including the masses of the quarks and leptons. We describe how to estimate the quark masses from consideration of spontaneous chiral symmetry.

A principal success of the electroweak theory was the prediction of *charm*. The $J/\psi$ particle discovered in 1974 was the first experimental signal of the charmed quark necessary to obtain a consistent description of neutral weak currents coupling to hadrons. The bottom and top quarks were subsequently discovered to fill out a third generation of quarks and leptons, with the $\tau$ lepton and its neutrino.

Precision data on the electroweak theory are now at the 0.1% level of accuracy and check the one-loop quantum corrections. A detailed comparison of many experimental results yields compelling support for the correctness of the theory. The one missing state is the Higgs boson, for which searches are under way at the large hadron collider at CERN.

Mixing of three quark generations gives rise to CP violation through a phase that could underlie the observed CP violation, although whether it explains the effect fully remains an open question. The strong CP problem remains a mystery, and potential difficulty, for the standard model.

## 4.2
## Lepton and Quark Masses

The elementary spin-$\frac{1}{2}$ fermions from which everything is made are taken to be the leptons and quarks. The leptons are singlets under the SU(3) color group and do not therefore experience strong interactions. Six flavors of lepton are known and these fall into the three doublets

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{4.1}$$

The masses of the charged leptons [1] are (in MeV) 0.511, 105.66, and 1783.5 for $e$, $\mu$, and $\tau$.

From big-bang cosmology, there are further interesting constraints on the number of neutrinos and on their masses. First, as pointed out by Gunn, Schramm, and Steigman [2, 3] the abundance of $^4$He observed in the universe and comparison with primordial nucleosynthesis indicates that the number of species of light ($\ll 1$ MeV) stable neutrinos should not exceed three. The reason is that each additional neutrino increases the energy density and hence the expansion rate so that the weak interactions converting protons to neutrons, and vice versa, freeze out earlier, to leave a larger neutron–proton ratio. Most of the neutrons end in $^4$He, so the primordial abundance of $^4$He is higher. The details are given in Refs. [2] and [3] and reviewed in Refs. [4] and [5]. Terrestrially, in 1989, the invisible width of the $Z$ gave the number of neutrinos with mass below $M_Z/2$ as 3.

Another cosmological constraint is an upper limit on the *sum* of the masses of neutrinos (i.e., light neutral stable particles). This limit is about 1 eV.

Quarks are triplets under $SU(3)_c$; antiquarks are antitriplets. Quarks appear in at least six flavors, which fall into the doublets

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix} \tag{4.2}$$

The electric charges will be taken as $+\frac{2}{3}$ and $-\frac{1}{3}$ for the upper and lower components, respectively. Integral charges have been suggested by Han and Nambu [13] and initially adapted in grand unification schemes by Pati and Salam [14–17], though without complete commitment [18]. For the electroweak interactions, only the average over colors matters, so we cannot distinguish the two possibilities; but the hadronic $e + e$-cross-section resolved this in favor of the fractional quark changes.

Until 1974 there was experimental evidence only for $u$, $d$, and $s$ which underlie the approximate flavor SU(3) symmetry of Gell-Mann and Ne'eman [19] (reprinted in Refs. [20–22]). The quarks were suggested as more than purely mathematical entities in 1964 by Gell-Mann [23] and Zweig [24] (reprinted in Ref. [25]). The acceptance of quarks as the true hadron constituents took a decade and received impetus in two important stages, in 1969 and 1974, respectively. The first (1969) was the observation of scaling in deep-inelastic electron–proton scattering, interpreted by Feynman [26] as due to pointlike constituents (partons) inside the proton. The identification of partons with quarks by Bjorken and Paschos [27, 28] led to the quark–parton model, which has proved indispensable phenomenologically (for reviews, see, e.g., Refs. [29] and [30]).

The second development (1974) involved the discovery of the $J/\psi$ particle [31, 32] and the subsequent appearance of both radial and orbital excitations, which confirmed that this hadron is an essentially nonrelativistic bound state of two spin-$\frac{1}{2}$ quarks: actually, a charmed quark and its antiparticle. The direct evidence for the occurrence of such heavy quarks obviously supports the idea that the light $u$,

$d$, and $s$ quarks be taken equally seriously as constituents of the light baryons and mesons.

A striking fact of particle physics is that although there is overwhelming evidence for quarks as hadron constituents, no evidence for free quarks exists (apart from some unreproduced exceptions; see, e.g., Refs. [33] and [34]). This confinement property is expected to be contained, as a consequence of infrared slavery, in QCD.

Quark confinement means, however, that specification of the quark masses is more ambiguous than for the lepton masses, and necessarily involves some arbitrariness in definition and convention. In particular, we must distinguish between constituent quark masses and current quark masses, the latter being most relevant for the QCD or electroweak Lagrangians used in perturbation theory. Here we follow the discussion by Weinberg [35].

Let us first focus on the three lightest quark flavors: $u$, $d$, and $s$. It is obvious from the hadronic mass spectrum that $s$ is somewhat more massive than $u$ and $d$. The masses can be defined and calculated most conveniently from the spontaneous breaking of $\mathrm{U}(3)_L \times \mathrm{U}(3)_R$ chiral symmetry [36, 37] in quantum chromodynamics. For massless quarks, the QCD Lagrangian (with quark color indices suppressed) is

$$L = \sum_k \bar{\psi}_k \gamma_\mu \left( \partial_\mu - \frac{1}{2} g \lambda^a A^a_\mu \right) \psi_k - \frac{1}{4} F^a_{\mu\nu} F^a_{\mu\nu} \tag{4.3}$$

This is invariant under the global chiral transformation ($a = 1, 2, \ldots, 9$)

$$\psi_{Lk} \to \left( e^{i\lambda^a \theta^a_L} \right)_{kl} \psi_{Ll} \tag{4.4}$$

$$\psi_{Rk} \to \left( e^{i\lambda^a \theta^a_R} \right)_{kl} \psi_{Rl} \tag{4.5}$$

where

$$\psi = \frac{1}{2}(1 + \gamma_5)\psi + \frac{1}{2}(1 - \gamma_5)\psi \tag{4.6}$$

$$= \psi_R + \psi_L \tag{4.7}$$

This invariance for the kinetic term is seen immediately since

$$\bar{\psi}\gamma_\mu\psi \equiv \bar{\psi}_R \gamma_\mu \psi_R + \bar{\psi}_L \gamma_\mu \psi_L \tag{4.8}$$

The "nonsimple" part $\mathrm{U}_L(1) \times \mathrm{U}_R(1) = \mathrm{U}_V(1) \times \mathrm{U}_A(1)$ corresponds to baryon number conservation $[\mathrm{U}_V(1)]$ and the axial $\mathrm{U}_A(1)$, which is known to be broken by nonperturbative instanton effects [38]. The relevant chiral symmetry is therefore only the semisimple $\mathrm{SU}(3)_V \times \mathrm{SU}(3)_A$ subgroup. Quark mass terms will break this chiral symmetry to only $\mathrm{SU}(3)_V$ since

$$m\bar{\psi}\psi \equiv m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R) \tag{4.9}$$

is invariant only if $\theta_L^a = \theta_R^a$ and if the quarks have equal nonzero mass $m_u = m_d \neq 0$. This flavor $\mathrm{SU}(3)_V$ is broken further to isospin $\mathrm{SU}(2)_V$ for $m_u = m_d \neq m_s$. If $\mathrm{SU}(3)_A$ is broken spontaneously, there are eight (approximately) massless pseudoscalar Nambu–Goldstone bosons, identified as the octet of pseudoscalar mesons ($\pi^\pm, \pi^\circ, K^\pm, K^\circ, \bar{K}^0, \eta$). We shall now indicate how using current algebra and PCAC, the meson masses may be related to the quark mass ratios provided that we assume that the vacuum is flavor SU(3) symmetric.

The PCAC relation

$$\partial_\mu J_\mu^{5a}(x) = F_\pi \mu_m^2 \left(\Pi^a\right) \phi^a(x) \tag{4.10}$$

gives for the pseudoscalar $\Pi^a$ propagator

$$E^{ab}(q) = \int d^4x e^{iq\cdot x} \langle 0| T\left(\partial_\mu J_\mu^{5a}(x), \partial_\nu J_\nu^{5b}(0)\right)|0\rangle \tag{4.11}$$

$$= \frac{i\delta^{ab} F_\pi^2 \mu_m^4(\Pi^a)}{\mu_m^2(\pi^a) - q^2} \tag{4.12}$$

Hence the Nambu–Goldstone boson mass is

$$\mu_m^2\left(\Pi^a\right) = -\frac{i}{F_\pi^2} E^{aa}(0) \tag{4.13}$$

But integration by parts gives

$$E^{ab}(q) = -iq_\mu \int d^4x e^{iq\cdot x} \langle 0| T\left(J_\mu^{5a}(x), \partial_\nu J_\nu^{5b}(0)\right)|0\rangle$$
$$- \int d^4x e^{iq\cdot x} \langle 0|\left[J_0^{5a}(x), \partial_\nu J_\nu^{5b}(0)\right]\delta(x_0)|0\rangle \tag{4.14}$$

where we used

$$T\left(J_\mu^{5a}(x), \partial_\nu J_\nu^{5b}(0)\right) = \theta(x_0) J_\mu^{5a}(x) \partial_\nu J_\nu^{5b}(0) + \theta(-x_0) \partial_\nu J_\nu^{5b}(0) J_\mu^{5a}(x) \tag{4.15}$$

and

$$\partial_\mu \theta(\pm x) = \pm \delta_{\mu 0} \delta(x_0) \tag{4.16}$$

We also have

$$\partial_\nu J_\nu^{5b}(0) = -i\left[H_m(0), \int d^3y J_0^{5b}(y)\right] \tag{4.17}$$

where $H_m$ is that part of the Hamiltonian density breaking the $\mathrm{SU}(3)_A$ symmetry, namely,

$$H_m = m_u \bar{u}u + n_d \bar{d}d + m_s \bar{s}s \tag{4.18}$$

Hence

$$\mu_m^2(\Pi^a) = -\frac{i}{F_\pi^2}\int d^4x d^4y \langle 0|\left[J_0^{5a}(x), \left[J_0^{5a}(y), H_m(0)\right]_{y_0=0}\right]_{x_0=0}|0\rangle \quad (4.19)$$

where $F_\pi$ is the charged pion decay constant, $F_\pi \simeq 190$ MeV. With the standard normalization of the Gell-Mann matrices [$\mathrm{Tr}(\lambda^a\lambda^b) = 2\delta^{ab}$], note that [here $q = (u, d, s)$]

$$H_m = \bar{q}\left[\frac{1}{2}(M_u - M_d)\lambda^3 + \frac{1}{2\sqrt{3}}(M_u + M_d - 2M_s)\lambda^8 + \frac{1}{\sqrt{6}}(M_u + M_d + M_s)\lambda^9\right]q \quad (4.20)$$

and

$$J_0^{5a} = -i\bar{q}\gamma_0\gamma_5\lambda^a q \quad (4.21)$$

Now using the algebra of SU(3) generators,

$$\left[\lambda^a, \lambda^b\right] = 2if^{abc}\lambda^c \quad (4.22)$$

we find that

$$\mu_m^2(\Pi^+) = \mu_m^2(\Pi^0) = \frac{4}{F_\pi^2}\left[M_u\langle\bar{u}u\rangle_0 + M_d\langle\bar{d}d\rangle_0\right] \quad (4.23)$$

$$\mu_m^2(K^+) = \frac{4}{F_\pi^2}\left[M_u\langle\bar{u}u\rangle_0 + M_d\langle\bar{s}s\rangle_0\right] \quad (4.24)$$

$$\mu_m^2(K^0) = \frac{4}{F_\pi^2}\left[M_u\langle\bar{d}d\rangle_0 + M_d\langle\bar{s}s\rangle_0\right] \quad (4.25)$$

To proceed further, we may assume that flavor SU(3) is not broken spontaneously, so that

$$\langle\bar{u}u\rangle_0 = \langle\bar{d}d\rangle_0 = \langle\bar{s}s\rangle_0 \quad (4.26)$$

More generally, we might assume that $\langle\bar{u}u\rangle_0 = \langle\bar{d}d\rangle_0$ (strong isospin) but $\langle\bar{s}s\rangle = R\langle\bar{d}d\rangle_0$. For the moment we take $R = 1$; the possibility $R \neq 1$ is discussed later.

The physical meson masses have a contribution from virtual photons, so that

$$\mu^2(\Pi^a) = \mu_m^2(\Pi^a) + \mu_\gamma^2(\Pi^a) \quad (4.27)$$

with

$$\mu_\gamma^2(\Pi^0) = \mu_\gamma^2(K^0) = 0 \quad (4.28)$$

$$\mu_\gamma^2(\Pi^+) = \mu_\gamma^2(K^+) \tag{4.29}$$

Taking this into account, straightforward algebra then leads to the quark mass ratios

$$\frac{M_d}{M_u} = \frac{(K^0) - (K^+) - (\pi^+)}{2(\pi^0) + (K^+) - (K^0) - (\pi^+)} \tag{4.30}$$

$$\frac{M_s}{M_d} = \frac{(K^0) + (K^+) - (\pi^+)}{(K^0) - (K^+) + (\pi^+)} \tag{4.31}$$

where we used the meson symbol to denote its squared mass.

Note that such squared meson masses are related to unsquared quark masses because of the difference between boson and fermion propagators. In units of $(\text{GeV})^2$, the observed masses are

$$\mu^2(K^+) = 0.2437\ \text{GeV}^2 \tag{4.32}$$

$$\mu^2(K^0) = 0.2477\ \text{GeV}^2 \tag{4.33}$$

$$\mu^2(\pi^+) = 0.0195\ \text{GeV}^2 \tag{4.34}$$

$$\mu^2(\pi^0) = 0.0182\ \text{GeV}^2 \tag{4.35}$$

and hence the numerical values of the quark mass ratios are

$$\frac{m_d}{m_u} = 1.8 \tag{4.36}$$

$$\frac{m_s}{m_d} = 20.1 \tag{4.37}$$

If we take $R \neq 1$, then $m_d/m_u$ is unchanged but $m_s/m_d = 20.1/R$; for example, if $R = 4$, the latter ratio is reduced to about 5.

Next, we consider the absolute values of the quark masses, and this requires a normalization prescription. A sensible procedure is to assume that

$$\langle \bar{q}_k q_k \rangle_h = N_{hk} Z_m \tag{4.38}$$

where $N_{hk}$ is the number of valence quarks of flavor $k$ in hadron $h$ and $m_k^*$ is defined as the renormalized quark mass by

$$m_k^* = z_m m_k \tag{4.39}$$

with $z_m$ a *universal* renormalization constant. Within a flavor SU(3) multiplet, the mass splittings are due mainly to $m_s^*$:

$$m_h \simeq m_0 + m_s^* N_{hs} \tag{4.40}$$

and hence $m_s^*$ can be estimated. One has the unitary multiplets (the pseudoscalars are unreliable here because of their special role as approximate Nambu–Goldstone bosons from spontaneous breaking of chiral symmetry) with mass splittings [1]:

1. Vector mesons:

$$m(\rho) = 776 \text{ MeV } (S = 0) \tag{4.41}$$

$$m(K^*) = 892 \text{ MeV } (S = -1) \tag{4.42}$$

2. Tensor mesons:

$$m(A_2) = 1312 \text{ MeV } (S = 0) \tag{4.43}$$

$$m(K^{**}) = 1434 \text{ MeV } (S = -1) \tag{4.44}$$

3. Octet baryons:

$$m(N) = 939 \text{ MeV } (S = 0) \tag{4.45}$$

$$m(\Xi) = 1318 \text{ MeV } (S = -2) \tag{4.46}$$

4. Decuplet baryons:

$$m(N^*) = 1232 \text{ MeV } (S = 0) \tag{4.47}$$

$$m(\Omega) = 1672 \text{ MeV } (S = -3) \tag{4.48}$$

These four multiplets give mean mass increments per unit strangeness of 116, 122, 190, and 147 MeV, respectively. As an approximate working average value, we may take

$$m_s^* = 150 \text{ MeV} \tag{4.49}$$

Using the ratios derived previously, we then obtain

$$m_u^* = 4.2 \text{ MeV} \tag{4.50}$$

$$m_d^* = 7.5 \text{ MeV} \tag{4.51}$$

Of course, these values are much less than the "constituent" quark masses, which may be taken as about $\frac{1}{2}m(\rho)$ or $\frac{1}{3}m(N)$, that is, 350 MeV for $u$ and $d$ and 500 MeV for $s$. Most of this mass comes not from mass terms in the QCD Lagrangian but from nonperturbative effects, necessary for infrared slavery and confinement, that are less amenable to estimation. If we assume that this extra contribution 350 MeV is flavor independent, we can then estimate the renormalized current masses for $c$, $b$, and $t$. Using the masses $m(J/\psi) = 3.10$ GeV and $m(T) = 9.4$ GeV, one arrives at

$$m_c^* = 1200 \text{ MeV} \tag{4.52}$$

$$m_b^* = 4400 \text{ MeV} \tag{4.53}$$

$$m_t^* = 174{,}000 \text{ MeV} \tag{4.54}$$

With respect to these masses, we should add the following remarks:

1. The different $(m_d^* - m_u^*)$ is not necessarily equal to the difference in the corresponding constituent quarks because of extra electromagnetic contributions (of order $\sim 350$ MeV $\times\ 1/137 \sim 2$ MeV); nevertheless, the value obtained by the combinations, for example,

$$\frac{1}{2}\left[m(n) - m(p) + m\left(\Xi^-\right) - m\left(\Xi^0\right)\right] = 3.7 \text{ MeV} \tag{4.55}$$

$$\frac{1}{2}\left[\left(\Sigma^-\right) - m\left(\Sigma^+\right)\right] = 4.0 \text{ MeV} \tag{4.56}$$

   are impressively close.
2. By increasing the value of $R$ from $R = 1$ to, say, $R = 4$, as an extreme case we increase $m_u^*$ and $m_d^*$ by a factor of 4 to, for example, $m_d^* = 30$ MeV. But then $(m_d^* - m_u^*) = 13.2$ MeV seems too high.
3. Other approaches to the quark masses, involving global fits to both the baryon spectrum and the baryon sigma term [39, 40] or exploiting the "MIT bag" model [41], generally give higher mass values. But it is futile to attempt a more accurate calculation than we have already given.

## 4.3 Weak Interactions of Quarks and Leptons

Electromagnetic interactions are described successfully by quantum electrodynamics (QED), which is an abelian U(1) gauge theory with coupling $g = e > 0$, the electric charge of the positron. Under a gauge transformation $\beta(x)$ the fields transform as

$$\phi_a \to e^{iq_a\beta(x)}\phi_a \tag{4.57}$$

where $q_a$ is the electric charge eigenvalue for $\phi_a$. For example, in the first family

$$e^- \to e^{-i\beta(x)}e^- \tag{4.58}$$

$$\nu_e \to \nu_e \tag{4.59}$$

$$u \to e^{2/3 i\beta(x)} u \tag{4.60}$$

$$d \to e^{-1/3 i\beta(x)} d \tag{4.61}$$

and similarly for $(\mu^-, \nu_\mu, c, s)$ and $(\tau^-, \nu_\tau, t, b)$.

The amplitude for the basic electromagnetic process is

$$-ieA_\mu J_\mu^{\mathrm{EM}} \tag{4.62}$$

with

$$J_\mu^{\mathrm{EM}} = -\bar{e}\gamma_\mu e + \frac{2}{3}\bar{u}\gamma_\mu u - \frac{1}{3}\bar{d}\gamma_\mu d + \cdots \tag{4.63}$$

Clearly, electric charge and all flavors are conserved. Because $J_\mu^{\mathrm{EM}}$ is nonchiral, parity is conserved; also, QED is even under $C$ and $T$ separately.

The electric charge U(1) is unbroken; hence the electromagnetic force is long ranged. There is a limit on the photon mass given by [42]

$$m_\gamma < 6 \times 10^{-16}\ \mathrm{eV} \tag{4.64}$$

obtained by space-probe measurements of the Jovian magnetosphere. The fine-structure constant is $\alpha_e = e^2/4\pi = (137.036)^{-1}$ at $q^2 = 0$.

Weak interactions are responsible for $\beta$-decay, $\mu$-decay, and most hyperon decays. These forces are very weak and have a range less than $10^{-2}$ fermi $= 10^{-15}$ cm, corresponding to an energy scale $>$ 20 GeV. The original knowledge of weak interactions (for a review of the subject up to 1969, see Ref. [43]; some of the important original papers are reprinted in Refs. [20], [44] and [45]) was based on the fact that strangeness, $S$, and the third component of strong isospin, $T_3$, are violated, although both are conserved by strong and electromagnetic interactions. Charge conjugation, C, and parity, P, are both violated, although the product CP is conserved approximately. There is a small violation of CP in the neutral kaon system which is a very important feature of the electroweak forces and is discussed in Section 4.7.

Prior to the discovery of neutral currents in 1973, most known properties of weak interactions were describable by a modernized Fermi [46, 47] (reprinted in Ref. [44]) interaction of the form

$$L_W = \frac{1}{2}\frac{G}{\sqrt{2}}\left(J_\mu J_\mu^\dagger + J_\mu^\dagger J_\mu\right) \tag{4.65}$$

with $G = 1.027 \times 10^{-5} M_p^{-2}$ ($M_p$ = proton mass). The charged weak current was given by Cabibbo [48] (see also Ref. [49]). The (V–A) structure had been established earlier in Refs. [50] and [51]; the conserved vector current hypothesis of Ref. [51] was discussed earlier in Ref. [52] as

$$J_\mu = \bar{e}\gamma_\mu(1-\gamma_5)\nu_e + \bar{\mu}\gamma_\mu(1-\gamma_5)\nu_\mu$$
$$+ \cos\theta_c \bar{d}\gamma_\mu(1-\gamma_5)u + \sin\theta_c \bar{s}\gamma_\mu(1-\gamma_5)u + \cdots \tag{4.66}$$

where only three flavors $(u, d, s)$ have been included, as well as only two families of leptons. Defining

$$\psi_L = \frac{1}{2}(1 - \gamma_5)\psi \tag{4.67}$$

allows one to rewrite

$$J_\mu = 2(\bar{e}_L\gamma_\mu\nu_{eL} + \bar{\mu}_L\gamma_\mu\nu_{\mu L} + \cos\theta_c\bar{d}_L\gamma_\mu u_L + \sin\theta_c\bar{s}_L\gamma_\mu u_L + \cdots) \tag{4.68}$$

in terms of two-component Weyl spinors. One finds phenomenologically the value of the Cabibbo angle $\sin\theta_c = 0.228$, so that rates for strangeness-changing decays have a suppression $\sim \tan^2\theta_c = 0.055$.

From Eq. (4.68) we see that charged current interactions satisfy $\Delta S = \Delta Q = \pm 1$. Processes with $\Delta Q = -\Delta S$ are never observed. In nonleptonic weak decays, the strong isospin satisfies the empirical selection rule $\Delta T = \frac{1}{2}$; this may be explicable in terms of, for example, $s(T = 0) \to u(T = \frac{1}{2})$, as discussed later in connection with CP violation.

Because the electroweak interactions violate C and P, it is useful to discuss here in detail how C and P act on fermions. Such a discussion facilitates the formulation of electroweak theory as well as preparing the ground for later treatment of grand unified theories. Consider a four-component Dirac field $\psi(\mathbf{x}, t)$ which annihilates a particle or creates an antiparticle. From it we may define two two-component Weyl spinors $\psi_{R,L}$ by

$$\psi_{R,L} = \frac{1}{2}(1 \pm \gamma_5)\psi \tag{4.69}$$

with adjoints $\bar{\psi} = \psi^+\gamma_0$ satisfying

$$\bar{\psi}_{R,L} = \bar{\psi} \cdot \frac{1}{2}(1 \mp \gamma_5) \tag{4.70}$$

so that the nonvanishing terms in fermion-conserving bilinears involve $LR$ and $RL$ for even numbers of Dirac matrices and $LL$ and $RR$ for odd numbers. For example,

$$\bar{\psi}\psi \equiv \bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R \tag{4.71}$$

$$\bar{\psi}\gamma_\mu\psi \equiv \bar{\psi}_R\gamma_\mu\psi_R + \bar{\psi}_L\gamma_\mu\psi_L \tag{4.72}$$

For zero rest mass, $\psi_{R(L)}$ annihilates a right (left)-handed neutrino or creates a left (right)-handed antineutrino. For nonzero mass $m$, the same becomes true only in the ultrarelativistic limit with energy $E \gg m$; nevertheless, the Weyl spinors can still be defined in this fully Lorentz invariant manner.

If neutrinos are precisely massless and only $\nu_L$ is ever observed together with $(\bar{\nu})_R$, only $\psi_L$ is needed to describe the physical degrees of freedom. This is the two-component neutrino theory [53–57]. The two-component description does not necessarily imply that the neutrino is massless, however, unless we also assume that there is a conserved fermion number. The point is that as an alternative to a Dirac mass term, Eq. (4.71), we may construct a Lorentz scalar

$$\epsilon_{\alpha\beta}\psi_L^\alpha\psi_L^\beta \tag{4.73}$$

from $\psi_L$ alone. Here, $\epsilon_{12} = -\epsilon_{21} = +1$ and $\alpha, \beta = 1, 2$ denote the two components of the Weyl spinor. Expression (4.73) is usually called a Majorana mass term and does not allow definition of a conserved fermion number. The presence of Majorana mass terms that violate lepton number is suggested by some grand unified theories of strong and electroweak interactions, as will be discussed much later.

We now digress to explain how expression (4.73) is a Lorentz scalar. (Some further details are given in, e.g., Ref. [12].) The spinors $\psi_L$ and $\psi_R$ need to be identified with the $(\frac{1}{2}, 0)$, $(0, \frac{1}{2})$ representations of the Lorentz group $O(4) \sim SU(2) \times SU(2)$. Because of the indefinite metric, although rotations may be identified by $\mathbf{J} = \frac{1}{2}\boldsymbol{\sigma}$, the boosts must be identified nonunitarily by $\mathbf{K} = \pm(i/2)\boldsymbol{\sigma}$ for $\psi_{RL}$, respectively. Thus the action of Lorentz transformations is given by

$$\psi_{R,L} \to \Lambda_{R,L}\psi_{R,L} \tag{4.74}$$

with

$$\Lambda_{R,L} = \exp\left[i\frac{\boldsymbol{\sigma}}{2}\cdot(\boldsymbol{\theta} \pm i\boldsymbol{\beta})\right] \tag{4.75}$$

Here $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are the rotation and boost parameters, respectively.

With the conventional choice for the Pauli matrices, one has

$$\sigma_2\sigma_i\sigma_2 = -\sigma_i^* \tag{4.76}$$

and hence

$$\sigma_2\Lambda_{R,L}^{-1}\sigma_2 = \Lambda_{R,L}^T \tag{4.77}$$

since under transposition

$$\sigma_i^T = \sigma_i^* \tag{4.78}$$

Thus we have

$$\lambda_{R,L}^T\sigma_2\Lambda_{R,L} = \sigma_2 \tag{4.79}$$

From Eq. (4.79) it follows that a Lorentz scalar bilinear in $\psi_R$ is given by

$$\psi_R^T\sigma_2\psi_R \to \psi_R^T\Lambda_R^T\sigma_2\lambda_R\psi_R \tag{4.80}$$

$$= \psi_R^T\sigma_2\psi_R \tag{4.81}$$

Since $\sigma_2$ is antisymmetric, this scalar is seen to arise from the antisymmetric part of $(0, \frac{1}{2}) \times (0, \frac{1}{2})$. Similarly, $\psi_L^T \sigma_2 \psi_L$ is a scalar, being the antisymmetric part of $(\frac{1}{2}, 0) \times (\frac{1}{2}, 0)$; this is then precisely of the form (up to a factor $-i$) of Eq. (4.73), as required.

Let us now consider the effects of P and C. Under P (parity), a Dirac spinor transforms as [see, e.g., Ref. [58], Eq. (2.33)]

$$\psi(\mathbf{x}, t) \xrightarrow{\mathrm{P}} \psi'(-\mathbf{x}, t) = \gamma_0 \psi(-\mathbf{x}, t) \tag{4.82}$$

because

$$p_\mu \xrightarrow{\mathrm{P}} p'_\mu = (p_{0'} - \mathbf{p}) \tag{4.83}$$

$$\not{p} = \gamma_0 \not{p}' \gamma_0 \tag{4.84}$$

$$(\not{p} - m)\psi(\mathbf{x}, t) = 0 \xrightarrow{\mathrm{P}} (\not{p}' - m)\psi'(-\mathbf{x}, t) = 0 \tag{4.85}$$

Note that there is an overall phase in Eq. (4.81) which has been chosen arbitrarily. It follows that for the Weyl spinors

$$\psi_{L,R}(\mathbf{x}, t) \xrightarrow{\mathrm{P}} \gamma_0 \psi_{R,L}(-\mathbf{x}, t) \tag{4.86}$$

and

$$\overline{\psi_{L,R}(\mathbf{x}, t)} \xrightarrow{\mathrm{P}} \overline{\psi_{R,L}(-\mathbf{x}, t)} \gamma_0 \tag{4.87}$$

The Cabibbo charged current involves terms of the form (not summed on $\mu$)

$$\overline{\psi_{1L}(\mathbf{x}, t)} \gamma_\mu \psi_{2L}(\mathbf{x}, t) \xrightarrow{\mathrm{P}} \overline{\psi_{1R}(-\mathbf{x}, t)} \gamma_\mu \psi_{2R}(-\mathbf{x}, t) g_{\mu\mu} \tag{4.88}$$

and clearly $L_w$ is not invariant under P. Note that $L_w$ involves $(g_{\mu\mu})^2 = +1$, so there is no problem with Lorentz invariance.

Under charge conjugation C, the behavior of a Dirac spinor is [cf. Ref. [58], Eq. (5.4)]

$$\psi \xrightarrow{\mathrm{C}} \psi^c = \mathrm{C}\bar{\psi}^T = \mathrm{C}\gamma_0 \psi^* \tag{4.89}$$

with (the overall phase is arbitrary)

$$\mathrm{C} = i\gamma_2\gamma_0 \tag{4.90}$$

This follows from requiring that

$$(i\not{\nabla} - e\not{A} - m)\psi = 0 \xrightarrow{\mathrm{C}} (i\not{\nabla} + e\not{A} - m)\psi^c = 0 \tag{4.91}$$

Hence

$$(\mathrm{C}\gamma^0)\gamma_\mu^*(\mathrm{C}\gamma^0)^{-1} = -\gamma_\mu \tag{4.92}$$

and since

$$\gamma^0 \gamma_\mu^* \gamma^0 = \gamma_\mu^T \tag{4.93}$$

one needs

$$\mathrm{C}\gamma_\mu^T \mathrm{C}^{-1} = -\gamma_\mu \tag{4.94}$$

In the standard representation with $\gamma_0$ and $\gamma_2$ symmetric and $\gamma_1$ and $\gamma_3$ antisymmetric, this mean that $[\mathrm{C}, \gamma_{0,2}] = \{\mathrm{C}, \gamma_{1,3}\}_+ = 0$; these relations are clearly satisfied by Eq. (4.90).

For the Weyl spinors, it follows that

$$\psi_{R,L} \xrightarrow{\mathrm{C}} (\psi_{R,L})^c = \left(\mathrm{C}\psi_{R,L}^*\right) = \left(\psi^c\right)_{L,R} \tag{4.95}$$

and hence in the charged weak current

$$\bar{\psi}_{1L}\gamma_\mu\psi_{2L} \xrightarrow{\mathrm{C}} -\bar{\psi}_{2R}\gamma_\mu\psi_{1R} \tag{4.96}$$

so that $L_w$ is *not* invariant.

Under the product CP, we have from Eqs. (4.88) and (4.96),

$$\bar{\psi}_{1L}(\mathbf{x}, t)\gamma_\mu\psi_{2L}(\mathbf{x}, t) \xrightarrow{\mathrm{CP}} \bar{\psi}_{2L}(-\mathbf{x}, t)\gamma_\mu\psi_{1L}(-\mathbf{x}, t) \tag{4.97}$$

and hence for the Lagrangian density $L_w$,

$$L_w(\mathbf{x}, t) \to L_w(-\mathbf{x}, t) \tag{4.98}$$

Thus the action

$$\int d^4x L_w(\mathbf{x}, t) \tag{4.99}$$

is invariant under CP. This assumed that the coupling G in Eq. (4.65) is real.

With the Cabibbo structure, $L_w$ described charged current weak interactions successfully, except CP violation and possibly the nonleptonic kaon and hyperon decays. However, $L_w$ cannot be exact since at high energies it violates unitarity. For example, in the process $\nu_e e^- \to \nu_e e^-$, the cross section is given by, ignoring masses,

$$\sigma(S) = \frac{G_F^2 S}{\pi} \tag{4.100}$$

where $S$ is the squared center of mass energy. But being a point interaction, the process is pure $S$-wave and hence must satisfy partial-wave unitarity of the total cross section:

$$\sigma(S) < \frac{4\pi}{S} \tag{4.101}$$

Thus if $S > 2\pi/G_F \sim (700\ \mathrm{GeV})^2$, there is a violation of unitarity.

We might hope to unitarize by higher-loop corrections, but the problem then is that the four-fermion interaction, Eq. (4.65), is nonrenormalizable and has uncontrollable ultraviolet divergences.

The next step is to assume that the four-fermion interaction is the low-energy limit of a finite-range force. The Lagrangian then becomes

$$L_w = \frac{g}{2\sqrt{2}}(J_\mu W_\mu^- + J_\mu^\dagger W_\mu^+) \tag{4.102}$$

with a propagator for the intermediate vector boson (IVB) of the form

$$D_{\mu\nu}(k) = \frac{-i(g_{\mu\nu} - k_\mu k_\nu / M_w^2)}{k^2 - M_w^2 + i\epsilon} \tag{4.103}$$

with the low-energy limit

$$D_{\mu\nu}(k) \underset{k^2 \ll M_w^2}{\longrightarrow} \frac{i g_{\mu\nu}}{M_w^2} \tag{4.104}$$

Hence with the identification

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_w^2} \tag{4.105}$$

the low-energy limit coincides for $L_w$ from Eqs. (4.65) and (4.102).

The IVB is like the photon of QED in being a vector (spin 1) intermediary, but there are very important differences: (1) the IVB is massive, (2) the IVB is electrically charged, and (3) the IVB couples only to left-handed particles and to right-handed antiparticles; that is, it violates parity conservation.

Introduction of the IVB partially cures the unitarity problem because $\nu_e e^- \to \nu_e e^-$ is no longer pure $S$-wave. But, for example, $e^+e^- \to W^+W^-$ violates unitarity for $\sqrt{S} \gtrsim G_F^{-1/2}$, as before. This is closely related to the nonrenormalizability arising from the $k_\mu k_\nu$ term in the propagator, Eq. (4.103), as detailed in Chapter 3. This sort of problem with unitarity afflicted weak interaction theory for a very long time from 1934 [46, 47] to 1971 [59, 60].

The problem is that the IVB is not yet a gauge theory, and for renormalizability we need to embed the Fermi theory somehow into a renormalizable quantum field theory. One suggestion was to make double scalar exchange mimic the V–A interaction [61, 62]. However, the standard model uses spontaneous symmetry breaking and a gauge theory.

In terms of the charged weak current $J_\mu(\mathbf{x}, t)$ defined in Eq. (4.66), define

$$Q^- = \int d^3x J_0(\mathbf{x}, t) \tag{4.106}$$

$$Q^+ = (Q^-)^\dagger \tag{4.107}$$

Also, the electromagnetic charge

$$Q = \int d^3x\, J_0^{\text{em}}(\mathbf{x}, t) \tag{4.108}$$

is a third generator. But then so is

$$\left[Q^-, Q^+\right] = 2\int d^3x\left[\bar{e}\gamma_0(1-\gamma_5)e - \bar{\nu}\gamma_0(1-\gamma_5)\nu + \cdots\right] \tag{4.109}$$

Hence, we need at least two neutral gauge bosons. One may try to avoid this, but the necessary Higgs structure is complicated [63].

Another possibility is to write a gauge field theory for the weak interactions alone [64], but the unification of more forces is clearly superior and actually correct (i.e., agrees with experiment).

Let us retrace the history of the standard model: The first article to propose a theory unifying weak and electromagnetic interactions was by Schwinger [65], who was unfortunately misled by incorrect experiments indicating tensor weak couplings. The principal ingredients of the now-standard SU(2) × U(1) electroweak theory were provided in 1960 by Glashow [66], who also surmised (see also Ref. [67]) that explicit symmetry breaking by vector mass terms might preserve renormalizability. A similar theory was proposed by Salam and Ward [68]. The Higgs mechanism [69–76] was introduced into SU(2) × U(1) theory by Weinberg [77], and Salam [78], normalized the neutral current strength; both conjectured that this softer spontaneous breaking would keep renormalizability. This Weinberg–Salam conjecture was vindicated a few years later through the work of 't Hooft [61], completed by him and Veltman [79, 80] and by Lee and Zinn-Justin [81].

As proposed initially (e.g., Ref. [77]), this theory included only the leptons; indeed, a straightforward extension to include the $u$, $d$, and $s$ flavors gave substantial strangeness-changing neutral currents, which disagreed with experimental limits. A crucial contribution in 1970 by Glashow, Iliopoulos, and Maiani (GIM) [82] showed how a fourth flavor, "charm" of quark, can cancel this unwanted neutral current. Combined with the GIM mechanism, the standard electroweak theory therefore stood in 1971 with two major predictions: (1) charm and (2) neutral currents, both of which were confirmed, as discussed in subsequent subsections.

Here we begin by introducing the SU(2) × U(1) gauge theory. The generators of the weak isospin and weak hypercharge are $T^i$ $(i = 1, 2, 3)$ and $Y$, respectively; the corresponding gauge vectors and couplings are $A^i_\mu$, $B_\mu$ and $g, g'$. The theory is a chiral one which violates parity by assigning $\psi_L$ and $\psi_R$ matter fields to different representations of SU(2) × U(1). All the $\psi_L$ are SU(2) doublets; all $\psi_R$ are singlets. Both $\psi_L$ and $\psi_R$ transform under U(1) such that electric charge $Q = T_3 + Y$. Thus, the left-handed quarks and leptons form SU(2) doublets:

$$q_{mL} = \begin{pmatrix} u_m \\ d_m \end{pmatrix}_L, \qquad l_{mL} = \begin{pmatrix} \nu_m, \\ e_m \end{pmatrix}_L \tag{4.110}$$

where $m = 1, 2, 3, \ldots$ is a family label. These doublets have $T_3 = \pm\frac{1}{2}$ and $Y = +\frac{1}{6}$ and $-\frac{1}{2}$, respectively; note that $Y$ is just the mean electric charge of the weak isomultiplet. The isosinglets $u_{mR}$, $d_{mR}$, and $e_{mR}$ have $Y = +\frac{2}{3}, -\frac{1}{3}$, and $-1$, respectively. The kinetic term for the fermions is therefore

$$L_f = \sum_m (\bar{q}_{mL} i \not{D} q_{mL} + \bar{l}_{mL} i \not{D} l_{mL} + \bar{u}_{mR} i \not{D} u_{mR} + \bar{d}_{mR} i \not{D} d_{mR} + \bar{e}_{mR} i \not{D} e_{mR}) \tag{4.111}$$

Here the covariant derivatives follow from the SU(2) × U(1) charges:

$$D_\mu q_{mL} = \left( \partial_\mu - i g \frac{\tau}{2} \cdot \mathbf{A}_\mu - i \frac{g'}{6} B_\mu \right) q_{mL} \tag{4.112}$$

$$D_\mu l_{mL} = \left( \partial_\mu - i g \frac{\tau}{2} \cdot \mathbf{A}_\mu - i \frac{g'}{6} B_\mu \right) l_{mL} \tag{4.113}$$

$$D_\mu u_{mR} = \left( \partial_\mu - i \frac{2}{3} g' B_\mu \right) u_{mR} \tag{4.114}$$

$$D_\mu d_{mR} = \left( \partial_\mu + \frac{i}{3} g' B_\mu \right) d_{mR} \tag{4.115}$$

$$D_\mu e_{mR} = (\partial_\mu + i g' B_\mu) e_{mR} \tag{4.116}$$

In the minimal model there is a Higgs scalar doublet ($Y = +\frac{1}{2}$)

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{4.117}$$

Hence the scalar piece $L_\phi$ is

$$L_\phi = \frac{1}{2} (D_\mu \phi)^\dagger (D_\mu \phi) - V(\phi) \tag{4.118}$$

with

$$D_\mu \phi = \left( \partial_\mu - i g \frac{\tau}{2} \cdot \mathbf{A}_\mu - i \frac{g'}{2} B_\mu \right) \phi \tag{4.119}$$

$$V(\phi) = m^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{4.120}$$

Here $V(\phi)$ is the most general SU(2) × U(1) invariant renormalizable potential.

The kinetic term for the gauge vectors is

$$L_V = -\frac{1}{4} F^i_{\mu\nu} F^i_{\mu\nu} - \frac{1}{4} G_{\mu\nu} G_{\mu\nu} \tag{4.121}$$

with

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g\epsilon^{ijk} A^j_\mu A^k_\nu \tag{4.122}$$

$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{4.123}$$

Finally, in the full classical Lagrangian $L$ there is a Yukawa piece $L_Y$ necessary to give the fermion masses upon spontaneous symmetry breaking:

$$L = L_f + L_\phi + L_V + L_Y \tag{4.124}$$

The explicit form of $L_Y$ is

$$L_Y = \sum_{m,n} \left( \Gamma^u_{mn} \bar{q}_{mL} \phi u_{nR} + \Gamma^d_{mn} \bar{q}_{mL} \phi d_{nR} + \Gamma^e_{mn} \bar{l}_{mL} \phi e_{nR} \right)$$

$$+ \text{Hermitian conjugate} \tag{4.125}$$

with a double sum over families. This then completes the unbroken classical Lagrangian.

## 4.4 Charm

The standard electroweak theory for leptons already predicts a weak neutral current with definite strength, isospin, and Lorentz properties. Incorporation of quarks needs another ingredient: the fourth, "charm" flavor of quark.

Let us first trace the history of the charm concept, since the evolution of this idea from 1962 through 1973 and the remarkable experimental results of November 1974 and the subsequent two years together form a beautiful paradigm of the "scientific method."

The first suggestion of a fourth flavor was motivated by an analogy between leptons and hadrons, or what in modern language would be the lepton–quark analogy. Before the postulation of quarks, such an analogy was based on the Sakata model [83], in which hadrons were supposed to be composed from the constituents $P$, $N$, and $\Lambda$. After the discovery [84] that $\nu_e$ and $\nu_\mu$ are different neutrinos, this analogy suggested [85–90] that a fourth constituent be considered for hadrons to keep the balance with the four leptons $e$, $\mu$, $\nu_e$, and $\nu_\mu$.

This argument was sharpened through the invention of flavor SU(3) by Gell-Mann [91, 92] (reprinted in Ref. [20]) and by Ne'eman [93] (reprinted in Ref. [20]). This SU(3) theory gained universal acceptance following the experimental discovery [94] (reprinted in Ref. [20]) of the predicted $\Omega^-$ particle. There followed the proposal by Gell-Mann [95] (reprinted in Ref. [20]) and Zweig [96] that there might be entities transforming as the defining representation of SU(3); these entities were named *quarks* by Gell-Mann [95]. Although in 1964 quarks were suggested as perhaps no more than mathematical quantities, they are now universally accepted as the actual constituents of hadrons, although they are probably permanently confined within hadrons and unable to exist as free particles. The two principal historical reasons for the promotion of quarks from mathematics to physics were (1) the experimental discovery in the late 1960s of scaling behavior in deep-inelastic electron–nucleon scattering which was interpreted as evidence for pointlike constituents in the nucleon, and (2) the experimental discovery in 1974 and subsequent years of hadrons containing the charmed quark, which is our present subject.

The success of flavor SU(3) as a generalization of SU(2) isospin led Tarjanne and Teplitz [97] to take the next step to SU(4); other authors [98–102] followed suit. Several of these papers again mention the quark–lepton analog but the general proposition is: If SU(3), why not SU(4)? In particular, Bjorken and Glashow [101] introduced the name *charm* to designate the fourth flavor of quark beyond up ($u$), down ($d$), and strange ($s$). Alternative schemes were proposed based, for example, on two triplets of quarks [103, 104], but such theories were not borne out by experiment.

Following this extensive theoretical activity providing somewhat vague motivations for charm, there followed a hiatus until 1970 when the true *raison d'etre* was given in the classic paper by Glashow, Iliopoulos, and Maiani [82], who introduced a mechanism for suppression of strangeness-changing weak neutral currents. This was incorporated into gauge field theories by Weinberg [105, 106]. Yet another motivation for charm is the cancellation of $\gamma_5$ triangle anomalies as emphasized by Bouchiat, Iliopoulos, and Meyer [107]. But the most physical argument was the GIM explanation [82] as to why the weak neutral current usually respects quark flavor (a fact noted much earlier [108]).

How does the fourth flavor help to suppress the strangeness-changing neutral current? Consider the electroweak forces alone acting on the doublets

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \quad \begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \tag{4.126}$$

and right-handed singlets

$$e_R^-, \quad \mu_R^-, \quad u_R, d_R, c_R, S_R \tag{4.127}$$

Suppose we first take all Higgs couplings to vanish so that there are massless fermions coupling to massless gauge bosons. In this limit, the charged gauge bosons $W_\mu^\pm$ couple only $u$ to $d$, and $c$ to $s$, within the same irreducible representation of SU(2), and no strangeness-changing processes such as $s \to uW^-$ exist. How, then, do the strangeness-changing processes $\Delta S = \Delta Q = \pm 1$ occur at all?

To obtain such processes, it is essential to break the symmetry by turning on the Higgs couplings, thus allowing the Higgs to develop nonzero vacuum expectation values (VEVs). The Higgs give masses to the quarks (and to the charged leptons), but there is no reason for the mass eigenstates thus generated simultaneously to be eigenstates of the gauge group; thus the Higgs mix $c \leftrightarrow u$ and $s \leftrightarrow d$ in expression (4.126). At first glance, this gives two Cabibbo angles, but one is unobservable since if we mix $c \to u$ and $s \to d$ by the same amount, nothing changes. The convention is to mix $d \leftrightarrow s$ so that expression (4.126) becomes

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L \quad \begin{pmatrix} c \\ s' \end{pmatrix}_L \tag{4.128}$$

$$d' = d \cos\theta_c + s \sin\theta_c \tag{4.129}$$

$$s' = -d \sin\theta_c + s \cos\theta_c \tag{4.130}$$

The neutral current is invariant under this rotation since

$$\bar{d}'d' + \bar{s}'s' = \bar{d}d + \bar{s}s \tag{4.131}$$

Thus the presence of $s'$, which was absent until motivated as a partner of charm, is the key to canceling the unwanted strangeness-changing neutral current. The original Cabibbo proposal would contain only $u$ and $d'$, but then if a gauge theory contained the charged gauge bosons $W^{\pm}_{\mu}$ coupling to generators of $SU(2)_L$ with the associated quark currents

$$J^{\mu}_{\pm} = \bar{q}\gamma^{\mu}\frac{1}{2}(1-\gamma_5)T^L_{\pm}q \tag{4.132}$$

it should also contain the neutral current

$$J^{\mu}_{0} = \bar{q}\gamma_{\mu}\frac{1}{2}(1-\gamma_5)T^L_{0}q \tag{4.133}$$

where $T^L_0 = [T^L_+, T^L_-]$. But with only $u$ and $d'$ one then finds that

$$J^{\mu}_{0} = \bar{s}\gamma_{\mu}\frac{1}{2}(1-\gamma_5)d \sin\theta_c \cos\theta_c \tag{4.134}$$

Such a strangeness-changing neutral weak current would disagree with the strong empirical bounds, such as

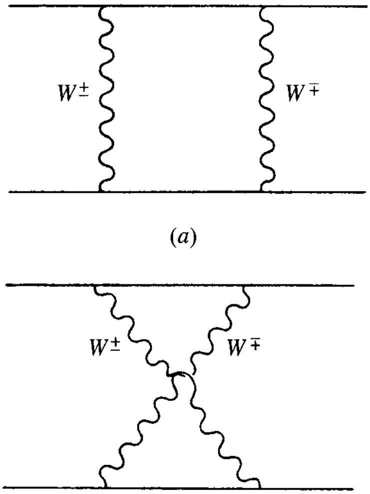
$$\frac{\Gamma(K^0_L \to \mu^+\mu^-)}{\Gamma(K^0_L \to \text{all})} \sim 10^{-8} \tag{4.135}$$

$$\frac{\Gamma(K^{\pm} \to \pi^{\pm}\nu\bar{\nu})}{\Gamma(K^{\pm} \to \text{all})} < 0.6 \times 10^{-6} \tag{4.136}$$

$$\frac{\Gamma(K^{\pm} \to \pi^{\pm}e^+e^-)}{\Gamma(K^{\pm} \to \text{all})} = (2.6 \pm 0.5) \times 10^{-7} \tag{4.137}$$

Once charm is added, however, the weak neutral current contains a piece

$$-\bar{s}\gamma^{\mu}\frac{1}{2}(1-\gamma_5)d \sin\theta_c \cos\theta_c \tag{4.138}$$




(b)

**Figure 4.1** Double $W^\pm$ exchange diagrams.

to cancel precisely Eq. (4.134) at lowest order, as well as giving substantial cancellations even at the one-loop level, as we shall see. Thus the mechanism gave a firm prediction (in 1970) of the existence of charm, which was discovered experimentally in 1974.

The mechanism can be generalized to any number of flavors if we postulate that the weak neutral current is flavor diagonal to order $G_\alpha$ [109–111]. We can change from mass eigenstates to flavor eigenstates by applying separate unitary transformations to the left and right helicities,

$$\not{q} = \frac{1}{2}(1-\gamma_5)U^L q + \frac{1}{2}(1+\gamma_5)U^R q \tag{4.139}$$

But then the neutral current coupling matrices are transformed according to

$$\mathscr{T}_0^{L,R} = U^{L,R} T_0^{L,R} \left(U^{L,R}\right)^{-1} \tag{4.140}$$

where $T_0^L$ is defined in Eq. (4.133) and $T_0^R$ would be nontrivial if there were also a gauge group under which the right-handed quark transformed. At order $G$, we require that $\mathscr{T}_0^{L,R}$ be flavor diagonal for any choice of $U^{L,R}$; this is the condition of naturalness that the property not depend on special tuning of parameters in the Lagrangian. The $U^{L,R}$ must commute with exactly conserved charges, such as the electric charge (or, of course, color). But then since

$$\left[U^{L,R}, Q\right] = 0 \tag{4.141}$$

the $T_0^{L,R}$ must be functions of quark charge: $T_0^{L,R}(Q)$. This accounts for the Born diagram. At order $G_\alpha$, there are further possible neutral interactions induced by double $W^\pm$ exchange (Fig. 4.1), as well as renormalizations of single $Z^0$ exchange

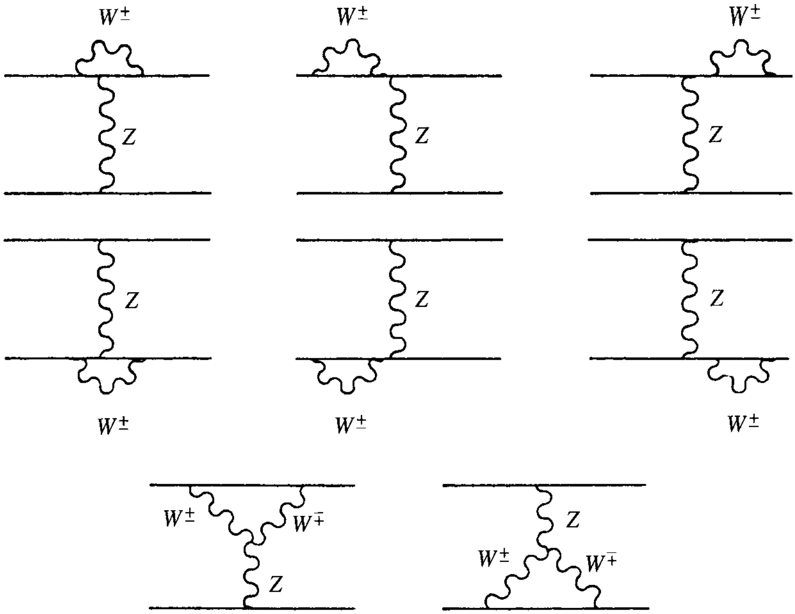



**Figure 4.2** Renormalization of single $Z^\circ$ exchange.

(Fig. 4.2). The latter are already diagonalized, but the diagrams of Fig. 4.1 give neutral couplings proportional to [110, 111]

$$\Delta T_0^{L,R} \sim 3\big[(\mathrm{T}^{L,R})^2 - (T_3^{L,R})^2\big] \pm 5(T_3^{L,R}) \tag{4.142}$$

where the $\pm$ signs refer to the uncrossed (Fig. 4.1*a*) and crossed (Fig. 4.1*b*) diagrams, respectively. Thus, to avoid flavor-changing neutral weak interactions at order $G_\alpha$ requires also that $(I^{L,R})^2 = (I^{L,R}(Q))^2$, a function only of charge.

To summarize: All quarks of the same charge must have the same value of $(I^{L,R})^2$ as well as of $(T_0^{L,R})$.

The mechanism for charm is a special case: Since $s$ has the same electric charge as $d$, it must have the same $T_3^L(= -\frac{1}{2})$ and the same $(I^L)^2 = \frac{1}{2}(\frac{1}{2}+1)$. Hence there must be a quark (c) with $T_3^L = +\frac{1}{2}$ in the same multiplet.

If we take the standard electroweak theory with gauge group SU(2) × U(1) and assume that *all* quarks have electric charge $\frac{2}{3}$ and $-\frac{1}{3}$, natural absence of flavor-changing neutral interactions dictates that all left-handed helicities are paired into SU(2) doublets, while all right-handed helicities are in SU(2) singlets. Hence discovery of the $b$ quarks with charge $-\frac{1}{3}$, as discussed below, implied the existence of a $t$ (top) quark with charge $+\frac{2}{3}$ in this sequential picture. It was possible to construct topless models (e.g. Refs. [112] and [113]) only by invoking a new quantum number for $b$ [112] or by maintaining GIM through a complicated Higgs sector [113]; of course, here we have been assuming weak processes to be mediated by vector gauge boson exchange, *not* by scalar exchanges.

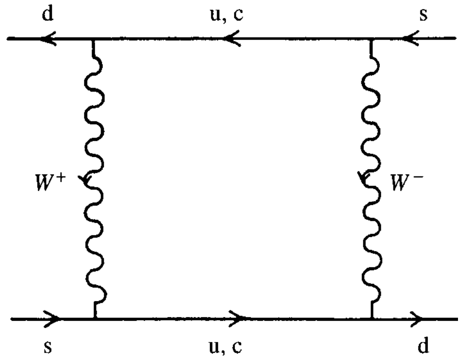



**Figure 4.3** Loop diagram for $K_1 - K_2$ mass difference.

Reverting to just four flavors, the next question is: Can the charmed quark have arbitrarily large mass and still cause the appropriate cancellation, or could the mass be predicted? The answer, presaged in reference [82] and provided in detail by Gaillard and Lee [114], is that the charm mass is highly constrained. The cancellation is precise in loops only if $m_c = m_u$ and hence the symmetry breaking is proportional to some power of the mass difference $(m_c - m_u)$; only if $m_c$ is not too large can one avoid disagreement with experiment. The most stringent limit is provided by the $K_1 - K_2$ mass difference. The relevant loop diagram is shown in Fig. 4.3 and gives an effective Lagrangian [114] of

$$L \sim -\frac{G_F}{\sqrt{2}}\frac{\alpha}{4\pi}\left(\frac{m_c}{37.6\ \text{GeV}}\right)^2 \cos^2\theta_c \sin^2\theta_c$$

$$\cdot\left[\bar{s}\gamma_\mu \frac{1}{2}(1-\gamma_5)d\bar{s}\gamma_\mu \frac{1}{2}(1-\gamma_5)d\right] \tag{4.143}$$

assuming that $m_w \gg m_c \gg m_u$. Here $\theta_c$ is the Cabibbo angle, $\alpha = (137.036)^{-1}$ and 37.6 GeV $= \frac{1}{2}ev$, $v$ being the VEV for the Higgs doublet. Taking the matrix element between $K_0$ and $\bar{K}_0$ states then gives for the mass difference $\Delta m_k$,

$$\Delta m_k \simeq \frac{1}{m_k}\langle\bar{K}_0| - L|K_0\rangle \tag{4.144}$$

$$= \frac{G_F}{\sqrt{2}} f_k^2 m_k \frac{\alpha}{4\pi}\left(\frac{m_c}{37.6\ \text{GeV}}\right)^2 \cos^2\theta_c \sin^2\theta_c \tag{4.145}$$

where $f_k$ is the charged kaon decay constant, $f_k \simeq 130$ MeV. Substituting $\sin\theta_c = 0.22$, $G_F m_p^2 = 1.02\times 10^{-5}$, and the measured value $\Delta m_k \simeq (0.7\times 10^{-14})m_k$ gives a prediction that the charmed quark mass is $m_c \simeq 1.5$–$2$ GeV. This mass is then consistent [114] with the other data on flavor-changing neutral currents given in Eqs. (4.135) through (4.137). Thus one might expect $(\bar{c}c)$ bound states to show up in the mass region 3 to 4 GeV. Also, the asymptotic freedom of QCD [115, 116] leads us to expect considerable weakening of the QCD coupling strength, and hence to a long lifetime if the $\bar{c}c$ state is lying below the threshold for decay into two charmed mesons. This prediction was available in 1974 but made slightly too late [117, 118] to anticipate the spectacular experimental discovery of November 1974. The long lifetime successfully exemplifies the quark-line rule suggested phenomenologically by Okubo, Zweig, and Iizuka [119–121] known commonly as the OZI rule.

The prediction of charm was verified dramatically by the simultaneous and independent discoveries at Brookhaven National Laboratory [122] and at Stanford Linear Accelerator Center [123] of a narrow vector resonance ($J/\psi$) at mass 3.10 GeV, and less than two weeks later SLAC [124] found a recurrence ($\psi'$) at 3.68 GeV. The striking appearance of these new states precipitated extensive theoretical analysis and further detailed experiments. The theory is reviewed in Ref. [125] and interprets $J/\psi$ as a *charmonium* system based on an atomic model of heavy $\bar{c}c$ quarks bound in a static, confining potential. Such a model is very successful in describing the spectrum of observed $\bar{c}c$ states (some representative early papers are Refs. [126–130]), The level spacings are much less than the mass scale, suggesting that the states may be treated nonrelativistically by the Schrödinger equation. This is not the whole story because the hyperfine (spin-dependent) splitting is comparable to the radial and orbital excitation energy. The type of static potential employed is typically

$$V(r) = -\frac{k}{r} + \frac{r}{a^2} \tag{4.146}$$

where $k \simeq 0.2$ and $1/a^2 \simeq 1\ \mathrm{GeV}/f \simeq 0.2\ (\mathrm{GeV})^2$.

These $c\bar{c}$ states have vanishing charm quantum number ($C = 0$) and it was in 1976 [131] that mesons with explicit charm were discovered. By now, the mesons necessary to fill out the (badly broken) flavor-SU(4) multiplets have been shown to exist for the vector and the pseudoscalar states. These developments confirm the reality of the charmed quark and hence indirectly that of the lighter $u$, $d$, and $s$ quarks. The mechanism requires that $c$ decay weakly more into $s$ ($\cos\theta_c$) that $d$ ($\sin\theta_c$). Indeed, experiment shows that, for example,

$$\frac{\Gamma(D^0 \to \pi^+\pi^-)}{\Gamma(D^0 \to K^-\pi^+)} \approx \tan^2\theta_c \tag{4.147}$$

thus confirming the expected couplings.

## 4.5 Bottom and Top Quarks

It was expected in 1975 that the discovery of charm might complete the flavor count of quarks. However, a third charged lepton, the $\tau$, was discovered at a mass of $M_\tau = 1777$ MeV [132, 133], signaling that the quark–lepton spectrum contained a third generation, since by quark–lepton symmetry one expected a corresponding third doublet of quarks.

This expectation was soon partially confirmed with the discovery of the bottom quark ($Q = -\frac{1}{3}$) [134] in 1977, first as a bottomonium bound state, the $\Upsilon$, and then as hadrons with explicit $B$ quantum number [135]. The mass of the bottom quark turned out to be $\sim 4.2$ GeV.

The top quark partner ($Q = \frac{2}{3}$) of the bottom quark proved elusive for well over a decade since it is remarkably heavier than all the other five quark flavors. It was eventually discovered in 1994 at Fermilab [136, 137] with a mass $\sim 174$ GeV.

Since experiments on the invisible $Z$ width strongly suggest only that three neutrinos exist with mass $< 45$ GeV, it appears likely that the top quark is the last flavor of light quark. See, however, Ref. [138] for a discussion of the possibility of further quark flavors beyond top.

## 4.6
## Precision Electroweak Data

As described in Chapter 7, the standard model (SM) has 19 free parameters. Nevertheless, its comparison to experimental data over the period 1973–2008 has given it more and more credence as the correct theory up to energies of at least $E \leqslant M_Z$. In particular, there have been extremely accurate (one per mille, or 1 in $10^3$) tests at the $Z$ factories of LEP at CERN and of SLC at SLAC.

The weak neutral currents predicted by the SM were discovered in 1973 by the Gargamelle experiment using the proton synchrotron (PS) at CERN [139] and then confirmed at Fermilab [140]. Subsequent neutrino scattering experiments confirmed the SM at the 1% level. The scattering of electrons from deuterons and protons, and atomic parity violation, being sensitive to $\gamma - Z$ interference, were also crucial to confirm the basic classical structure of the SM. The $W^{\pm}$ and $Z^0$ were eventually discovered in 1982 and 1983 by the UA1 [141] and UA2 [142, 143] groups at CERN, respectively. The history of these developments is given in Ref. [144].

The three colliders involved in the higher precision (0.1%) tests were SLC (the SLD detector), the LEP collaborations (ALEPH, DELPHI, OPAL, and Z3) and the Tevatron at Fermilab (CDF and DO collaborations). These colliders derived exceptionally accurate (0.002%) determination of the $Z$ mass and one-per-mille accuracy for numerous other observables which are tabulated in this section. These measurements were made at the $Z$-pole in $e^+e^-$ scattering and are based on over 20 million such $Z$ events.

It is necessary to define the observables before presenting the experimental comparison to the SM. The total width of the $Z$ is $\Gamma_Z$ and partial decay widths are $\Gamma(\text{had})$ for hadronic decays, $\Gamma(\text{inv})$ for invisible decays and $\Gamma(l^+l^-)$ for decays $l = e, \mu, \tau$, which should be equal from universality and neglecting the lepton masses. Thus $\Gamma_Z = \Gamma(\text{had}) + \Gamma(\text{inv}) + 3\Gamma(l^+l^-)$ within the SM. The cross section $\sigma_{\text{had}}$ refers to the process $e^+e^- \to Z^0 \to$ hadrons at the resonance peak.

**Table 4.1** $Z$-Pole Precision Observables from LEP and the SLC

| Quantity | Group(s) | Experimental results | Standard model[a)] | Pull |
|---|---|---|---|---|
| $M_z$(GeV) | LEP | $91.1867 \pm 0.0021$ | $91.1865 \pm 0.0021$ | 0.1 |
| $T_z$(GeV) | LEP | $2.4939 \pm 0.0024$ | $2.4957 \pm 0.0017$ | −0.8 |
| $\Gamma$(had)(GeV) | LEP | $1.7423 \pm 0.0023$ | $1.7424 \pm 0.0016$ | 0.0 |
| $\Gamma$(inv)(MeV) | LEP | $500.1 \pm 1.9$ | $501.6 \pm 0.2$ | 0.0 |
| $\Gamma(l^+l^-)$(MeV) | LEP | $83.90 \pm 0.10$ | $83.98 \pm 0.03$ | 0.0 |
| $\sigma_{\text{had}}(nb)$ | LEP | $41.491 \pm 0.058$ | $41.473 \pm 0.015$ | 0.3 |
| $R_e$ | LEP | $20.783 \pm 0.052$ | $20.748 \pm 0.019$ | 0.7 |
| $R_\mu$ | LEP | $20.789 \pm 0.034$ | $20.749 \pm 0.019$ | 1.2 |
| $R_\tau$ | LEP | $20.764 \pm 0.045$ | $20.794 \pm 0.019$ | −0.7 |
| $A_{\text{FB}}(e)$ | LEP | $0.0153 \pm 0.0025$ | $0.0161 \pm 0.003$ | −0.3 |
| $A_{\text{FB}}(\mu)$ | LEP | $0.0164 \pm 0.0013$ | | 0.2 |
| $A_{\text{FB}}(\tau)$ | LEP | $0.0183 \pm 0.0017$ | | 1.3 |
| $R_b$ | LEP + SLD | $0.21656 \pm 0.00074$ | $0.2158 \pm 0.0002$ | 1.0 |
| $R_c$ | LEP + SLD | $0.1735 \pm 0.0044$ | $0.1723 \pm 0.0001$ | 0.3 |
| $R_{s,d}/R_{u+d+s}$ | Opal | $0.371 \pm 0.023$ | $0.3592 \pm 0.0001$ | 0.5 |
| $A_{\text{FB}}(b)$ | LEP | $0.0990 \pm 0.0021$ | $0.1028 \pm 0.0010$ | −1.8 |
| $A_{\text{FB}}(c)$ | LEP | $0.0709 \pm 0.0044$ | $0.0734 \pm 0.0008$ | −0.6 |
| $A_{\text{FB}}(s)$ | Delphi + Opal | $0.101 \pm 0.015$ | $0.1029 \pm 0.0010$ | −1.0 |
| $A_b$ | SLD | $0.867 \pm 0.035$ | $0.9347 \pm 0.0001$ | −1.9 |
| $A_c$ | SLD | $0.647 \pm 0.040$ | $0.6676 \pm 0.0006$ | −0.5 |
| $A_s$ | SLD | $0.82 \pm 0.12$ | $0.9356 \pm 0.0001$ | −1.0 |
| $A_{LR}$(hadrons) | SLD | $0.1510 \pm 0.0025$ | $0.1466 \pm 0.0015$ | 1.8 |

*Source:* Data from Ref. [145].

a) The SM errors are from uncertainties in $M_Z$, ln $M_H$, $\alpha(M_Z)$, and $\alpha_S$. They have been treated as Gaussian; correlations are incorporated.

The $R_l$ are the ratios of the leptonic-to-hadronic widths $R_l = \Gamma(l^+l^-)/\Gamma(\text{had})$ for $l = e, \mu, \tau$. The $A_{\text{FB}}(l)$ are the forward–backward asymmetries $A_{\text{FB}}(l) = \frac{3}{4}A_e A_l$. In this relationship

$$A_f \frac{2v_f a_f}{v_f^2 + a_f^2} \tag{4.148}$$

with $a_f = I_{3,f}$ and $v_f = I_{3,f} - 2Q_f \bar{s}_f^2$ as the axial-vector and vector $Zf\bar{f}$ couplings. The effective electroweak mixing angle is written $\bar{s}_f^2 = \sin^2 \Theta_f^{\text{eff}}$.

In Table 4.1 are shown comparisons of numerous observables between the SM and experiment. The last column gives the "pull", which is the number of standard deviations by which the experimental value differs relative to the SM prediction. The overall fit is an impressive success for the SM since no pull is even as much as two standard deviations.

At higher energy above the $Z$-pole, LEP2 [146], together with the CDF [147] and DO [148] collaborations at Fermilab, have determined the $W$ mass from $W^+W^-$

production as

$$M_W = 80.388 \pm 0.063 \text{ GeV} \tag{4.149}$$

The top quark has been discovered in 1994 and its mass has an overall mean value

$$m_t = 173.8 \pm 3.2(\text{stat}) \pm 3.9(\text{syst}) \text{ GeV} \tag{4.150}$$

Lower-energy data in deep-inelastic scattering, atomic parity violation, and $(g-2)_\mu$ are all fully consistent with the standard model. Deviations from the SM are conveniently parameterized by oblique parameters. The first of these to be defined historically [149] was

$$\rho_0 = \frac{M_W^2}{M_Z^2 c_Z^2 \hat{\rho}(m_t, M_H)} \tag{4.151}$$

where $\hat{\rho}$ includes the SM contributions and hence $\rho = 1$, by definition. Related to $\rho_0$ is the quantity $T$ [150]

$$T = \frac{1 - \rho^{-1}}{\alpha} \tag{4.152}$$

The $S$ parameter is defined through

$$\frac{\alpha(M_Z)S}{4s_Z^2 c_Z^2} = \frac{\Pi_{ZZ}^{\text{new}}(M_Z^2) - \Pi_{ZZ}^{\text{new}}(0)}{M_Z^2} \tag{4.153}$$

where the "new" superscript means to include only new physics contributions. An overall fit gives [145]

$$S = -0.27 \pm 0.12 \tag{4.154}$$

$$T = 0.00 \pm 0.15 \tag{4.155}$$

These ranges of $S$ and $T$ are a severe constraint on the forms of new physics allowed. For example, although low-energy supersymmetry is allowed, most technicolor models give too positive a value for $S$ and are excluded.

## 4.7 Higgs Boson

The only particle in the standard model remaining to be discovered is the Higgs boson. In the minimal standard the expected mass range is [145]

$$114.4 \text{ GeV} < M_H < 194 \text{ GeV} \tag{4.156}$$

where the lower bound is a direct experimental limit from the combined LEP experiments at the four detectors ALEPH, DELPHI, L3 and OPAL.

The upper bound in (4.156) comes from comparison of radiative corrections to the standard model with precision electroweak data. The entire range of Higgs mass allowed by (4.156) would be accessible to the LHC expected to be commissioned at CERN in 2008 with pp collisions at center of mass energy 14 TeV and design luminosity $10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

Of course, once discovered, it would be useful to construct a Higgs factory: for example, a muon collider running continuously at the direct-channel Higgs boson resonance.

## 4.8 Quark Flavor Mixing and CP Violation

The gauge group of the standard model is $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ broken at the weak scale to $\mathrm{SU}(3)_C \times \mathrm{U}(1)_Y$. Under the standard group the first generation transforms as

$$Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \bar{u}_L, \bar{d}_L; \qquad L_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, e_L^+ \tag{4.157}$$

and the second $(c, s, \nu_\mu, \mu)$ and third $(t, b, \nu_\tau, \tau)$ generations are assigned similarly.

Quarks acquire mass from the vacuum expectation value (VEV) of a complex $\mathrm{SU}(2)_L$ doublet of scalar $\phi = \binom{\phi^+}{\phi^0}$, giving rise to up and down quark mass matrices:

$$M(U) = \lambda_{ij}^U \langle\phi^0\rangle; \qquad M(D) = \lambda_{ij}^D \langle\phi^0\rangle \tag{4.158}$$

which are arbitrary matrices that may, without loss of generality, be chosen to be Hermitian. The matrices $M(U)$ and $M(D)$ of Eq. (4.158) are defined so that the Yukawa terms give, for example, $\bar{Q}_L M(U) u_R$ + hermitian conjugate and can be diagonalized by a biunitary transformation:

$$K(U)_L M(U) K(U)_R^{-1} = \mathrm{diag}(m_u, m_c, m_t) \tag{4.159}$$

$$K(D)_L M(D) K(D)_R^{-1} = \mathrm{diag}(m_d, m_s, m_b) \tag{4.160}$$

These mass eigenstates do not coincide with the gauge eigenstates of Eq. (4.157) and hence the charged $W$ couple to the left-handed mass eigenstates through the $3 \times 3$ CKM (Cabibbo–Kobayashi–Maskawa) matrix $V_{\mathrm{CKM}}$, defined by

$$V_{\mathrm{CKM}} = K(U)_L K(D)_L^{-1} \tag{4.161}$$

This is a 3 × 3 unitary matrix that would in general have nine real parameters. However, the five relative phases of the six quark flavors can be removed to leave just four parameters, comprising three mixing angles and a phase. This KM phase underlies the KM mechanism of CP violation.

With $N$ generations and hence an $N \times N$ mixing matrix there are $N(N-1)/2$ mixing angles and $(N-1)^2$ parameters in the generalized CKM matrix. The number of CP violating phases is therefore $(N-1)^2 - N(N-1)/2 = (N-1)(N-2)/2$. This is zero for $N = 2$, one for $N = 3$, three for $N = 4$, and so on. In particular, as Kobayashi and Maskawa [151] pointed out, with three generations there is automatically this source of CP violation arising from the 3 × 3 mixing matrix. This is the most conservative approach to CP violation. This source of CP violation is necessarily present in the standard model; the only question was whether it is the *only* source of CP violation. When the only observation of CP violation remained in the neutral kaon system, there was not yet sufficient experimental data to answer this question definitively. With the data from B factories, a positive answer emerged.

There are various equivalent ways of parameterizing the CKM matrix. That proposed [151] by KM involved writing

$$V_{\text{CKM}} = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1\cos\theta_3 & -\sin\theta_1\sin\theta_3 \\ \sin\theta_1\cos\theta_2 & \cos\theta_1\cos\theta_2\cos\theta_3 - \sin\theta_2\sin\theta_3 e^{i\delta} & \cos\theta_1\cos\theta_2\sin\theta_3 + \sin\theta_2\cos\theta_3 e^{i\delta} \\ \sin\theta_1\sin\theta_2 & \cos\theta_1\sin\theta_2\cos\theta_3 + \cos\theta_2\sin\theta_3 e^{i\delta} & \cos\theta_1\sin\theta_2\sin\theta_3 - \cos\theta_2\cos\theta_3 e^{i\delta} \end{pmatrix} \tag{4.162}$$

Another useful parametrization [152] writes

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & \lambda^2 A \\ \lambda^3 A(1 - \rho - \eta) & -\lambda^2 A & 1 \end{pmatrix} \tag{4.163}$$

In Eq. (4.163), $\lambda$ is the sine of the Cabibbo angle $\sin\theta_1$ in Eq. (4.162) and CP violation is proportional to $\eta$. If we write the CKM matrix a third time as

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{4.164}$$

then the unitarity equation $(V_{\text{CKM}})^\dagger V_{\text{CKM}} = 1$ dictates, for example, that

$$V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0 \tag{4.165}$$

This relation is conveniently represented as the addition of three Argand vectors to zero is a *unitarity triangle*. Dividing out the middle term of Eq. (4.165) and using the parametrization of Eq. (4.163) leads to the prediction of the standard model with KM mechanism that the vertices of the unitarity triangle in a $\rho$–$\eta$ plot are at the origin $(0, 0)$, at $(1, 0)$, and at $(\rho, \eta)$. Thus, the area of the unitarity triangle is proportional to $\eta$ and hence to the amount of CP violation. The measurement of the angles and sides of this unitarity triangle were the principal goals of the $B$ factories (see, e.g., Ref. [153] for a review).

Next we turn to a brief outline of the strong CP problem in the standard model. (More detailed reviews are available in Refs. [154–156].) The starting observation is that one may add to the QCD Lagrangian an extra term:

$$L = \sum_k \bar{q}_k (i\gamma_\mu d_\mu - m) q_k - \Theta G_{\mu\nu} \mathcal{G}_{\mu\nu} \tag{4.166}$$

where the sum over $k$ is for the quark flavors and $D_\mu$ is the partial derivative for gauged color $SU(3)_C$. The additional term proportional to $\Theta$ violates P and CP symmetries. This term is a total divergence of a gauge noninvariant current but can contribute because of the existence of classical instanton solutions. It turns out that chiral transformations can change the value of $\Theta$ via the color anomaly but cannot change the combination:

$$\bar{\Theta} = \Theta - \arg\det M(U) - \arg\det M(D) \tag{4.167}$$

where $\det M(U, D)$ are the determinants of the up and down quark mass matrices, respectively. Thus $\bar{\Theta}$, which is an invariant under chiral transformations, measures the violation of CP symmetry by strong interactions. A severe constraint on $\bar{\Theta}$ arises from the neutron electric dipole moment $d_n$, which has been measured to obey $d_n \leqslant 10^{-25} e - \text{cm}$ [157, 158]. A calculation of $d_n$ [159, 160] leads to an estimate that $\bar{\Theta} < 10^{-10}$. This fine-tuning of $\bar{\Theta}$ is unexplained by the unadorned standard model and raises a serious difficulty thereto.

A popular approach (which does *not* necessitate additional fermions) involves the axion mechanism, which we describe briefly, although since only a relatively narrow window remains for the axion mass, and since the mechanism is non unique, it is well worth looking for alternatives to the axion for solving the strong CP problem.

In the axion approach, one introduces a color-anomalous global U(1) Peccei–Quinn symmetry [161, 162] such that different Higgs doublets couple to the up and down quarks. The effective potential now becomes a function of the two Higgs fields and $\bar{\Theta}(x)$ regarded as a dynamical variable. An analysis then shows that the potential acquires the form

$$V = V(H_1, H_2) - \cos\bar{\Theta} \tag{4.168}$$

and hence the minimum energy condition relaxes $\bar{\Theta}$ to zero.

Because a continuous global symmetry is broken spontaneously, there is a pseudo-Goldstone boson [163, 164] the axion, which acquires a mass through the color anomaly and instanton effects. The simplest model predicts an axion with mass of a few times 100 keV, but this particle was ruled out phenomenologically. Extensions of the axion model [165–168] lead to an axion mass which becomes a free parameter. Empirics constrain the mass to lie between about a micro-electron volt and a milli-electron volt, and searches are under way for such an axion.

In the kaon system, the CP violation parameter $\epsilon'/\epsilon$, which measures *direct* CP violation in the decay amplitude $K_L \to 3\pi$ (as opposed to indirect CP violation in $K^0 - \bar{K}^0$ mixing),

$$\frac{\epsilon'}{\epsilon} = \frac{1}{6}\left(\frac{[K_L \to \pi^+\pi^-]/[K_S \to \pi^+\pi^-]}{[K_L \to \pi^0\pi^0]/[K_s \to \pi^0\pi^0]} - 1\right) \tag{4.169}$$

The value found by KTeV [175] at Fermilab and by NA48 [176] at CERN, both in 1999, averages to the value

$$\frac{\epsilon'}{\epsilon} = 2.1 \times 10^{-3} \tag{4.170}$$

One other limit on CP violation is from the neutron electric dipole moment, for which the present limit is $d_n \leqslant 6 \times 10^{-26}$ electron-cm. Observation of nonzero $d_n$ and evaluation of the parameters $\rho$ and $\eta$ in Eq. (4.163) from B decay suggest strongly that the KM mechanism is responsible for all of the observed CP violation.

## 4.9 Summary

We have seen how the precision measurements, especially at the $Z$-pole, brilliantly confirm the standard electroweak theory. The Higgs boson remains undiscovered and the most enigmatic of the standard model particles: Is it really an elementary scalar field or a composite? This should be clarified at the LHC.

Flavor mixing between three families gives a natural mechanism for CP violation and it remains to be checked empirically whether this is the only source for CP noninvariance. The strong CP problem is an outstanding difficulty for QCD, as is the proliferation of parameters in the standard theory.

Going beyond the standard model, for example to grand unification, gives hope that some of the many parameters may be related, but generally it adds more and gives rise to the new problem of the "hierarchy": the huge ratio of the GUT scale to the weak scale. It is possible that a completely different approach to extending the SM will be necessary.

## References


1 Particle Data Group, *Eur. Phys. J.* C3, 1 (1998).
2 G. Steigman, D.N. Schramm, and J.E. Gunn, *Phys. Lett.* 66B, 202 (1977).
3 J. Yang, D.N. Schramm, G. Steigman, and R.T. Rood, *Astrophysical J.* 227, 697 (1979).
4 G. Steigman, *Annu. Rev. Nucl. Part. Sci.* 29, 313 (1979).
5 G. Steigman, in *Proceedings of the First Workshop on Grand Unification*, P.H. Frampton, S.L. Glashow, and A. Yildiz, eds. Math. Sci. Press, Brookline, MA, 1980, p. 245.
6 S.S. Gerstein and Ya.B. Zeldovich, *Zh. Eksp. Teor. Fiz. Pis'ma Red.* 4, 174 (1966).
7 R. Cowsik and J. McClelland, *Phys. Rev. Lett.* 29, 669 (1972)

**8** A.D. Dolgov and Ya.B. Zeldovich, *Rev. Mod. Phys.* 53, 1 (1980).

**9** SuperKamiokande Collaboration (Y. Fukuda et al.), *Phys. Rev. Lett.* 82, 2644 (1999).

**10** SuperKamiokande Collaboration (Y. Fukuda et al.), *Phys. Lett.* B467, 185 (1999).

**11** SuperKamiokande Collaboration (Y. Fukuda et al.), *Phys. Lett.* B452, 418 (1999).

**12** E. Kearas, T. Kajita, and Y. Totsuka, *Sci. Am.* 281, No. 2, 48 (1999).

**13** M.Y. Han and Y. Nambu, *Phys. Rev.* 139B, 1006 (1965).

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

**15** J.C. Pati and A. Salam, *Phys. Rev. Lett.* 31, 661 (1973).

**16** J.C. Pati and A. Salam, *Phys. Rev.*, D10, 275 (1974).

**17** J.C. Pati and A. Salam, *Phys. Lett.* 58B, 333 (1975).

**18** J.C. Pati and A. Salam, in *Proceedings of the First Workshop on Grand Unification*, P.H. Frampton, S.L. Glashow, and A. Yildiz, eds., Math. Sci. Press, Brookline, MA, 1980, p. 115.

**19** M. Gell-Mann, Caltech preprint, unpublished, 1961.

**20** M. Gell-Mann and Y. Ne'eman, *The Eightfold Way*, W.A. Benjamin, New York, 1964.

**21** Y. Ne'eman, *Nucl. Phys.* 26, 222 (1961).

**22** M. Gell-Mann, *Phys. Rev.* 125, 1067 (1962).

**23** M. Gell-Mann, *Phys. Lett.* 8, 214 (1964).

**24** G. Zweig, CERN Report 8419/TH412, unpublished, Developments in the Quark Model of Hadrons, Vol. I, 1964–1978.

**25** D.B. Lichtenberg and S.P. Rosen, eds., Hadronic Press, Cambridge, MA, 1980, p. 22.

**26** R.P. Feynman, *Phys. Rev. Lett.* 23, 1415 (1969).

**27** J.D. Bjorken and E.A. Paschos, *Phys. Rev.* 185, 1975 (1969).

**28** J.D. Bjorken and E.A. Paschos, *Phys. Rev.* D1, 3151 (1970).

**29** F.E. Close, *Introduction to Quarks and Partons*, Academic Press, New York, 1978.

**30** O.W. Greenberg, *Annu. Rev. Nucl. Part. Sci.* 28, 327 (1978).

**31** J.J. Aubert et al., *Phys. Rev. Lett.* 33, 1404 (1974).

**32** J.E. Augustin et al., *Phys. Rev. Lett.* 33, 1406 (1974).

**33** G.S. LaRue, W.M. Fairbank, and A.F. Hebard, *Phys. Rev. Lett.* 38, 1011 (1977).

**34** G.S. LaRue, W.M. Fairbank, and J.D. Phillips, *Phys. Rev. Lett.* 42, 142 (1979).

**35** S. Weinberg, *Trans. N.Y. Acad. Sci.* 38, 185 (1977) (Festschrift for I.I. Rabi).

**36** S.L. Glashow and S. Weinberg, *Phys. Rev. Lett.* 20, 224 (1968).

**37** M. Gell-Mann, R.J. Oakes, and B. Renner, *Phys. Rev.* 175, 2195 (1968).

**38** G. 't Hooft, *Phys. Rev. Lett.* 37, 8 (1976).

**39** J.F. Gunion, P.C. McNamee, and M.D. Scadron, *Nucl. Phys.* B123, 445 (1977).

**40** N.H. Fuchs and M.D. Scadron, *Phys. Rev.* D20, 2421 (1979).

**41** E. Golowich, *Phys. Rev.* D12, 2108 (1975).

**42** L. Davis, A.S. Goldhaber, and M.M. Nieto, *Phys. Rev. Lett.* 35, 1402 (1975).

**43** R.E. Marshak, Riazzudin, and C.P. Ryan, *Theory of Weak Interaction of Elementary Particles*, Wiley–Interscience, New York, 1969.

**44** P.K. Kabir, ed., *The Development of Weak Interaction Theory*, Gordon and Breach, New York, 1963.

**45** S.L. Adler and R.F. Dashen, *Current Algebras and Applications to Particle Physics*, W.A. Benjamin, New York, 1968.

**46** E. Fermi, *Nuovo Cimento* 11, 1 (1934).

**47** E. Fermi, *Z. Phys.* 88, 161 (1934).

**48** N. Cabibbo, *Phys. Rev. Lett.* 10, 531 (1963).

**49** M. Gell-Mann and M. Levy, *Nuovo Cimento* 16, 705 (1960).

**50** E.C.G. Sudarshan and R.E. Marshak, *Phys. Rev.* 109, 1860 (1958).

**51** R.P. Feynman and M. Gell-Mann, *Phys. Rev.* 109, 193 (1958).

**52** S.S. Gershtein and Ya.B. Zeldovich, Zh. Eks. *Teor. Fiz.* 29, 698 (1955) [trans.: *Sov. Phys. JETP* 2, 596 (1956)].

**53** H. Weyl, *Z. Phys.* 56, 330 (1929).

**54** A. Salam, *Nuovo Cimento* 5, 299 (1957).

**55** A. Salam, *Phys. Rev. Lett.* 2, 127 (1959).

**56** L. Landau, *Nucl. Phys.* 3, 127 (1957).

**57** T.D. Lee and C.N. Yang, *Phys. Rev.* 105, 1671 (1957).

**58** J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics*, McGraw–Hill, New York, 1964.
**59** G. 't Hooft, *Nucl. Phys.* B33, 173 (1971).
**60** G. 't Hooft, *Nucl. Phys.* B35, 167 (1971).
**61** W. Kummer and G. Segre, *Nucl. Phys.* 64, 585 (1965).
**62** D.A. Dicus, G. Segre, and V.L. Teplitz, *Phys. Rev.* D13, 3092 (1976).
**63** H. Georgi and S.L. Glashow, *Phys. Rev. Lett.* 28, 1494 (1972).
**64** S.A. Bludman, *Nuovo Cimento* 9, 433 (1958).
**65** J. Schwinger, *Ann. Phys.* 2, 407 (1957).
**66** S.L. Glashow, *Nucl. Phys.* 22, 579 (1961).
**67** S.L. Glashow, *Nucl. Phys.* 10, 107 (1959).
**68** A. Salam and J.C. Ward, *Phys. Lett.* 13, 168 (1964).
**69** P.W. Anderson, *Phys. Rev.* 130, 439 (1963).
**70** P.W. Higgs, *Phys. Lett.* 12, 132 (1964).
**71** P.W. Higgs, *Phys. Rev. Lett.* 13, 508 (1964).
**72** P.W. Higgs, *Phys. Rev.* 145, 1156 (1966).
**73** G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble, *Phys. Rev. Lett.* 13, 585 (1964).
**74** T.W.B. Kibble, *Phys. Rev.* 155, 1554 (1967).
**75** F. Englert and R. Brout, *Phys. Rev. Lett.* 13, 321 (1964).
**76** F. Englert, R. Brout, and M.F. Thiry, *Nuovo Cimento* 48, 244 (1966).
**77** S. Weinberg, *Phys. Rev. Lett.* 27, 1264 (1967).
**78** A. Salam, in *Elementary Particle Theory, Proceedings of the 8th Nobel Symposium*, N. Svartholm, ed., Almqvist & Wiksell, Stockholm, 1968.
**79** G. 't Hooft and M. Veltman, *Nucl. Phys.* B44, 189 (1972).
**80** G. 't Hooft and M. Veltman, *Nucl. Phys.* B50, 318 (1972).
**81** B. W. Lee and J. Zinn-Justin, *Phys. Rev.* D5, 3121, 3137, 3155 (1972).
**82** S.L. Glashow, J. Iliopoulos, and L. Maiani, *Phys. Rev.* D2, 1285 (1970).
**83** S. Sakata, *Prog. Theor. Phys.* 16, 686 (1956).
**84** G. Danby, J.M. Gaillard, K. Goulianos, L.M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger, *Phys. Rev. Lett.* 9, 36 (1962).
**85** Y. Katayama, K. Matumoto, S. Tanaka, and E. Yamada, *Prog. Theor. Phys.* 28, 675 (1962).
**86** Z. Maki, M. Nakagawa, and S. Sakata, *Prog. Theor. Phys.* 28, 870 (1962).
**87** Z. Maki, *Prog. Theor. Phys.* 31, 331 (1964).
**88** Z. Maki, *Prog. Theor. Phys.* 31, 333 (1964).
**89** Y. Hara, *Phys. Rev.* 134, B701 (1964).
**90** Z. Maki and Y. Ohnuki, *Prog. Theor. Phys.* 32, 144 (1964).
**91** M. Gell-Mann, unpublished, 1961.
**92** M. Gell-Mann, *Phys. Rev.* 125, 1067 (1962).
**93** Y. Ne'eman, *Nucl. Phys.* 26, 222 (1961).
**94** V.E. Barnes et al., *Phys. Rev. Lett.* 12, 204 (1964).
**95** M. Gell-Mann, *Phys. Lett.* 8, 214 (1964).
**96** G. Zweig, unpublished, 1964.
**97** P. Tarjanne and V.L. Teplitz, *Phys. Rev. Lett.* 11, 447 (1963).
**98** I. Sogami, *Prog. Theor. Phys.* 31, 725 (1964).
**99** M. Kama, K. Matumoto, and S. Tanaka, *Prog. Theor. Phys.* 31, 1171 (1964).
**100** D. Amati, H. Bacry, J. Nuyts, and J. Prentki, *Phys. Lett.* 11, 190 (1964).
**101** B.J. Bjorken and S.L. Glashow, *Phys. Lett.* 11, 255 (1964).
**102** D. Amati, H. Bacry, J. Nuyts, and J. Prentki, *Nuovo Cimento* 34, 1732 (1964).
**103** H. Bacry, J. Nuyts, and L. Van Hove, *Phys. Lett.* 9, 279 (1964).
**104** H. Bacry, J. Nuyts, and L. Van Hove, *Nuovo Cimento* 35, 510 (1964).
**105** S. Weinberg, *Phys. Rev. Lett.* 27, 1688 (1971).
**106** S. Weinberg, *Phys. Rev.* D5, 1412 (1972).
**107** C. Bouchiat, J. Iliopoulos, and Ph. Meyer, *Phys. Lett.* 38B, 519 (1972).
**108** M. Gell-Mann and S.L. Glashow, *Ann. Phys.* 15, 437 (1961).
**109** K. Kang and J.E. Kim, *Phys. Lett.* 64B, 93 (1976).
**110** S.L. Glashow and S. Weinberg, *Phys. Rev.* D15, 1958 (1977).
**111** E.A. Paschos, *Phys. Rev.* D15, 1966 (1977).
**112** H. Georgi and S.L. Glashow, *Nucl. Phys.* B167, 173 (1980).
**113** P.H. Cox, P.H. Frampton, and A. Yildiz, *Phys. Rev. Lett.* 46, 1051 (1981).
**114** M.K. Gaillard and B.W. Lee, *Phys. Rev.* D10, 897 (1974).

**115** D.J. Gross and F. Wilczek, *Phys. Rev. Lett.* 26, 1343 (1973).
**116** H.D. Politzer, *Phys. Rev. Lett.* 26, 1346 (1973).
**117** T. Appelquist and H.D. Politzer, *Phys. Rev. Lett.* 34, 43 (1975).
**118** T. Appelquist and H.D. Politzer, *Phys. Rev.* D12, 1404 (1975).
**119** S. Okubo, *Phys. Lett.* 5, 165 (1963).
**120** G. Zweig, unpublished, 1964.
**121** J. Iizuka, *Suppl. Prog. Theor. Phys.* 37/38, 21 (1966).
**122** J.J. Aubert et al., *Phys. Rev. Lett.* 33, 1404 (1974).
**123** J.E. Augustin et al., *Phys. Rev. Lett.* 33, 1406 (1974).
**124** G.S. Abrams et al., *Phys. Rev. Lett.* 33, 1453 (1974).
**125** T. Appelquist, R.M. Barnett, and K. Lane, *Annu. Rev. Nucl. Part. Sci.* 28, 387 (1978).
**126** T. Appelquist, A. De Rujula, H.D. Politzer, and S.L. Glashow, *Phys. Rev. Lett.* 34, 365 (1975).
**127** E. Eichten, K. Gottfried, T. Kinoshita, J. Kogut, K.D. Lane, and T.M. Yan, *Phys. Rev. Lett.* 34, 369 (1975).
**128** E. Eichten, K. Gottfried, T. Kinoshita, K.D. Lane, and T.M. Yan, *Phys. Rev. Lett.* 36, 500 (1976).
**129** B.J. Harrington, S.Y. Park, and A. Yildiz, *Phys. Rev. Lett.* 34, 168 (1975).
**130** B.J. Harrington, S.Y. Park, and A. Yildiz, *Phys. Rev. Lett.* 34, 706 (1975).
**131** G. Goldhaber et al., *Phys. Rev. Lett.* 37, 255 (1976).
**132** M. Perl et al., *Phys. Rev. Lett.* 35, 1489 (1975).
**133** R. Brandelik et al., *Phys. Lett.* 73B, 109 (1978).
**134** S.W. Herb et al., *Phys. Rev. Lett.* 39, 252 (1977).
**135** S. Behrends et al., *Phys. Rev. Lett.* 50, 881 (1983).
**136** R. Abe et al. (CDF collaboration), *Phys. Rev. Lett.* 74, 2626 (1995).
**137** S. Abachi et al. (D0 collaboration), *Phys. Rev. Lett.* 74, 2632 (1995).
**138** P.H. Frampton, P.Q. Hung, and M. Sher, *Phys. Rep.* 330, 263 (2000).
**139** F.J. Hasert et al. (Gargamelle collaboration), *Phys. Lett.* 46B, 138 (1973).
**140** A. Benvenuti et al. (HPWF collaboration), *Phys. Rev. Lett.* 32, 1457 (1974).
**141** G. Arnison et al. (UA1 collaboration), *Phys. Lett.* 122B, 103 (1983).
**142** M. Banner et al. (UA2 collaboration), *Phys. Lett.* 122B, 103 (1983).
**143** M. Banner et al. (UA2 collaboration), *Phys. Lett.* 126B, 398 (1983).
**144** A.K. Mann and D.B. Cline, eds., *Discovery of Weak Neutral Currents: The Weak Interaction Before and After, AIP Conference Proceedings*, 300, AIP Publishing, New York, 1994.
**145** W.-M. Yao et al., *J. Phys.* G33, 1 (2006).
**146** D. Karlen, talk presented at ICHEP'98.
**147** T. Dorigo, for the CDF collaboration, talk presented at the 6th International Symposium on Particles, Strings, and Cosmology, Boston, March 1998.
**148** B. Abbott et al., *Phys. Rev. Lett.* 80, 3008 (1998).
**149** M. Veltman, *Nucl. Phys.* B123, 89 (1977).
**150** M.E. Peskin and T. Takeuchi, *Phys. Rev. Lett.* 65, 964 (1990).
**151** M. Kobayashi and T. Mashkawa, *Prog. Theor. Phys.* 49, 652 (1973).
**152** L. Wolfenstein, *Phys. Rev. Lett.* 51, 1945 (1983).
**153** Y. Nir and H.R. Quinn, *Annu. Rev. Nucl. Part. Sci.* 42, 211 (1992).
**154** J.E. Kim, *Phys. Rep.* 150, 1 (1987).
**155** H.-Y. Cheng, *Phys. Rep.* 158, 1 (1988).
**156** R.D. Peccei, in *CP Violation*, C. Jarlskog, ed., World Scientific, Singapore, 1989.
**157** N.F. Ramsey, *Annu. Rev. Nucl. Sci.* 32, 211 (1982).
**158** K.F. Smith et al., *Phys. Lett.* 234, 191 (1990).
**159** V. Baluni, *Phys. Rev.* D19, 2227 (1979).
**160** R.J. Crewther, P. Di Vecchia, G. Veneziano, and E. Witten, *Phys. Lett.* 88B, 123 (1979).
**161** R.D. Peccei and H.R. Quinn, *Phys. Rev. Lett.* 38, 1440 (1977).
**162** R.D. Peccei and H.R. Quinn, *Phys. Rev.* D16, 1791 (1977).
**163** S. Weinberg, *Phys. Rev. Lett.* 40, 223 (1978).
**164** F. Wilczek, *Phys. Rev. Lett.* 40, 279 (1978).
**165** J.E. Kim, *Phys. Rev. Lett.* 43, 103 (1979).
**166** M.A. Shifman, A.I. Veinshtein, and V.I. Zhakarov, *Nucl. Phys.* B166, 493 (1980).
**167** A.P. Zhitnitskii, *Sov. J. Nucl. Phys.* 31, 260 (1980).
**168** M. Dine, W. Fischler, and M. Srednicki, *Phys. Lett.* 104B, 199 (1981).

**169** J. Gasser and H. Leutwyler, *Phys. Rep*. 87, 77 (1982).

**170** H. Leutwyler, *Nucl. Phys*. B337, 108 (1990).

**171** H. Leutwyler, *Phys. Lett*. B374, 163 (1996).

**172** H. Leutwyler, *Phys. Lett*. B378, 313 (1996).

**173** P.H. Frampton and T.W. Kephart, *Phys. Rev. Lett*. 66, 1666 (1991).

**174** P.H. Frampton and D. Ng, *Phys. Rev.* D43, 3034 (1991).

**175** A. Alavi-Harati et al., *Phys. Rev. Lett*. 83, 22 (1999).

**176** S. Palestini (NA48 collaboration), *International Europhysics Conference on High Energy Physics (EPS-HEP-99)*, Tampere, Finland, July 1999, hep-ex/9909046.

# 5
# Renormalization Group

## 5.1
## Introduction

Here we introduce the renormalization group equations, which are essential for understanding of quantum chromodynamics (QCD) and grand unification. This approach was introduced into quantum electrodynamics by Gell-Mann and Low in 1954, then remained somewhat in the background in elementary particle theory until the 1970s.

The discovery of asymptotic freedom of non-Abelian gauge theories by 't Hooft in 1972 played a major role in establishing QCD as the generally accepted theory of strong interactions between quarks, since it provided a natural explanation for why quarks become weakly coupled at very short distances. Also, it enables us to explore the short-distance regime by perturbative expansion in a small coupling constant.

Shortly afterward, in 1974, it was realized by Georgi, Quinn, and Weinberg that by extrapolating the renormalization group equations to even much shorter distances, one could speculate on how the strong and electroweak interactions might unify into a single coupling constant at about $10^{-29}$ cm. This improved the understanding of grand unified theories [e.g., the SU(5) theory of Georgi and Glashow], there remains no empirical support for this level of unification.

We first derive the Callan–Symanzik equation, which characterizes the renormalization group. Then the $\beta^{-}$ function for Yang–Mills theory will be calculated in covariant gauge; there follows the conditions for asymptotic freedom. Next, grand unification of the strong and electroweak interactions is discussed.

To illustrate use of anomalous dimensions, we analyze scaling violations in deep inelastic lepton–hadron scattering and show the quantitative agreement between QCD and experiment. Finally, the background field gauge is explained. This method makes calculations much easier and, consequently, gained popularity.

## 5.2 Renormalization Group Equations

The renormalization group involves the manipulation of a deceptively simple differential equation leading to surprisingly strong results concerning, for example, the asymptotic behavior of Green's functions for large external momenta; the results apply to the sum of all Feynman diagrams, including every order of perturbation theory. For the differential equation to be useful it is necessary that the quantum field theory have (1) ultraviolet divergences and (2) renormalizability.

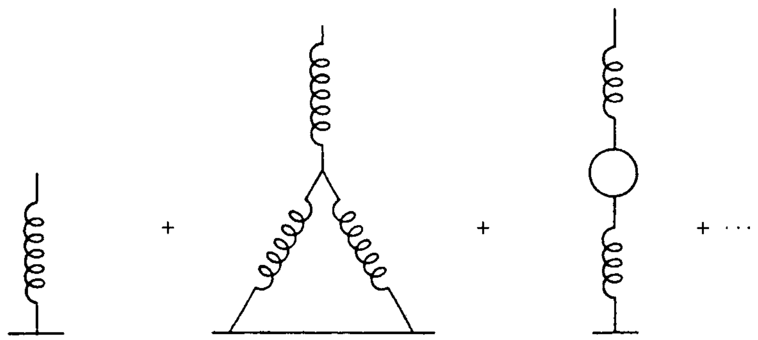
The renormalization group has a history that goes back to 1953 with the work of Stückelberg and Petermann [1], Gell-Mann and Low [2], and Ovsiannikov [3]. A good summary of the results of the 1950s is provided by Bogoliubov and Shirkov [4, Chap. 8], who also provide citations of the relevant Russian papers.

After more than a decade of quiescence (in particle theory, although not in condensed matter theory; see, e.g., Ref. [5]), the experimental observation of approximate scaling in deep-inelastic electron–nucleon scattering prompted a resurgence of interest in the renormalization group, in particular the important contributions of Callan [6, 7] and Symanzik [8–10] in 1970. Also of significance was the introduction of the operator product expansion by Wilson in 1968 [11].

The first great success of the renormalization group in particle theory, specifically in gauge field theories, was the discovery by 't Hooft [12], subsequently emphasized by Politzer [13, 14], and Gross and Wilczek [15–17] that QCD is asymptotically free and hence can explain the approximate scaling observed experimentally. A Yang–Mills non-Abelian gauge theory is essential to achieve this result [18, 19]. The asymptotic freedom was, and still is, a dominant reason for the general acceptance of QCD as the correct theory of strong interactions. A second use for the renormalization group came in 1974, when Georgi, Quinn, and Weinberg [20] demonstrated how strong and electroweak couplings may be unified, hence supporting the SU(5) model of Georgi and Glashow [21]. Excellent reviews of the renormalization group in field theory exist [14, 22–28].

Let us first consider the QCD of quarks and gluons. The full quark–gluon vertex, which includes all connected (and not only the proper or one-particle irreducible) Feynman diagrams will have contributions of order $g, g^3, g^5, \ldots$ in the renormalized perturbation expansion, corresponding to diagrams such as those in Fig. 5.1. Let $q_\mu$ be the momentum of the outgoing gluon–Euclidean spacelike (nonexceptional) $-q^2 \gg \mu^2$, where $\mu^2 > 0$ is arbitrary but preassigned. Then one finds that the leading order-by-order behavior of the full vertex is

$$g \to g + O\left(g^3 \ln\left(\frac{-q^2}{\mu^2}\right)\right) + O\left(g^5 \ln^2\left(\frac{-q^2}{\mu^2}\right)\right) + \cdots \tag{5.1}$$



(a) (b) (c)

**Figure 5.1** Corrections to quark–gluon vertex.

and these leading logarithms actually sum to

$$g^2(q^2) = \frac{g^2}{1 + bg^2 \ln(-q^2/\mu^2)} \tag{5.2}$$

The fact that Eq. (5.1) is a geometric series in $g^2 \ln(-q^2/\mu^2)$ would be impossible to verify by direct calculation of Feynman diagrams but will be an elementary consequence of the renormalization group equation.

Consider the proper Green's function $\Gamma^{(n)}(p_1, p_2, \ldots, p_n)$ with $n$ external lines and with naive dimension $\Gamma^{(n)} \sim M^d$ with $d$ = integer. In the limit $p_i^2 \to \infty$ with the $p_i p_j / \sqrt{p_i^2 p_j^2}$ held finite, we would at tree level obtain canonical scaling in terms of such kinematic invariants (having overall dimension $d$) provided that the $(-p_i^2)$ all greatly exceed every squared physical mass parameter in the theory. In the quantum theory, however, this is no longer true since at the *regularization* stage an arbitrary scale $\mu$ must be introduced: It is the mass unit of $\Lambda$ in Pauli–Villars regularization, and it is the corresponding scale in dimensional regularization where $(4 - n)^{-1} \leftrightarrow \ln(\Lambda/\mu)$. Note that in dimensional regularization we must introduce $\mu$ in order to relate the pole in the complex dimension plane to the logarithmic divergence of the loop-momentum integral, merely on dimensional grounds [$d$ is dimensionless but $\Lambda$ is not]. Thus, in the renormalized Green's function we expect that a change in $\mu$ can be compensated by a change in $g$ and the renormalization constants $Z$. If $n = n_A + n_F$, where $n_A$ and $n_F$ are the numbers of external gluons and quarks, respectively, the renormalizability of QCD enables us to write

$$\Gamma(p_i, \mu, g) = \lim_{\Lambda \to \infty} \left[ Z_A \left( \frac{g_0, \Lambda}{\mu} \right)^{n_A} Z_F \left( \frac{g_0, \Lambda}{\mu} \right)^{n_F} \Gamma_u^{(n)}(p_i, g_0, \Lambda) \right] \tag{5.3}$$

where $\Gamma_u^{(n)}$ is the unrenormalized proper Green's function and $g_0$ is the bare coupling constant.

The key observation now is that $\Gamma_u^{(n)}$ simply does not depend on $\mu$ since $\mu$ is introduced only by the regularization procedure. Operate on both sides of Eq. (5.3) with $\mu d/d\mu$ to find

$$\left(\mu\frac{\partial}{\partial\mu} + \mu\frac{\partial g}{\partial\mu}\frac{\partial}{\partial g}\right)\Gamma(p_i, \mu, g)$$
$$= \mu \lim_{\Lambda\to\infty}\left(\frac{n_A}{Z_A}\frac{\partial Z_A}{\partial\mu} + \frac{n_F}{Z_F}\frac{\partial Z_F}{\partial\mu}\right) Z_A^{n_A} Z_F^{n_F} \Gamma_u^{(n)}(p_i, g_0, \Lambda) \tag{5.4}$$

Hence if we make the definitions

$$\beta(g) = \lim_{\Lambda\to\infty}\left[\mu\frac{\partial}{\partial\mu} g\left(\frac{g_0, \Lambda}{\mu}\right)\right] \tag{5.5}$$

$$\gamma_i(g) = -\lim_{\Lambda\to\infty}\left[\mu\frac{\partial}{\partial\mu}\ln Z_i\left(\frac{g_0, \Lambda}{\mu}\right)\right] \tag{5.6}$$

where $i = A, F$, we arrive at the fundamental equation

$$\left[\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g} + n_A\gamma_A(g) + n_F\gamma_F(g)\right]\Gamma^{(n_A,n_F)}(p_i, \mu, g) = 0 \tag{5.7}$$

This first-order partial differential equation in two variables $\mu$ and $g$ is the renormalization group equation. Here $\beta(g)$ is the Callan–Symanzik $\beta$-function or, henceforth, simply the $\beta$-function, while the $\gamma_i(g)$ are, for reasons to become clear shortly, the anomalous dimensions. We may say that while renormalization is itself technically complicated but conceptually simple, the renormalization group is technically simple but conceptually subtle. For example Eq. (5.7) yields extraordinarily powerful results for partial sums (e.g., leading logarithms) over all orders of perturbation theory.

Instead of varying $\mu$ with $p_i$ fixed, we wish to make $p_i \to \lambda p_i$, but this is easy since $\mu$ also sets the scale for $p_i$. We may always write

$$\Gamma^{(n_A,n_F)}(\lambda p_i, \mu, g) = \mu^d f\left(\frac{\lambda^2 p_i p_j}{\mu^2}\right) \tag{5.8}$$

where $d$, as before, is the naive dimension and $f$ is a dimensionless function. Putting $t = \ln\lambda$, one finds immediately that

$$\left(\mu\frac{\partial}{\partial\mu} + \frac{\partial}{\partial t} - d\right)\Gamma^{(n_A,n_F)}(\lambda p_i, \mu, g) = 0 \tag{5.9}$$

and, hence, substitution into Eq. (5.7) gives

$$\left[-\frac{\partial}{\partial t} + \beta(g)\frac{\partial}{\partial g} + \gamma(g)\right]\Gamma^{(n_A,n_F)}(\lambda p_i, g) = 0 \tag{5.10}$$

with

$$\gamma(g) = d + n_A\gamma_A(g) + n_F\gamma_F(g) \tag{5.11}$$

This explains why the $\gamma_i(g)$ are named anomalous dimensions, since if $\beta(g) = 0$, we see that Eq. (5.10) implies that

$$\Gamma^{(n_A,n_F)} \sim \lambda^{d+n_A\gamma_A+n_F\gamma_F} \tag{5.12}$$

as $\lambda \to \infty$, so that scale invariance holds but with the naive dimension $d$ modified by the $\gamma_i(g)$.

When $\beta(g) \neq 0$, we must solve Eq. (5.10). The two independent variables $g$ and $t = \ln\lambda$ make this tricky to visualize at first, but a beautiful mechanical analog makes the solution perspicuous [22]. Consider the equation for $\rho(x, t)$:

$$\frac{\partial\rho}{\partial t} + v(x)\frac{\partial\rho}{\partial x} = L(x)\rho \tag{5.13}$$

This describes the density p of bacteria moving in a fluid in a pipe where $v(x)$ is the fluid velocity and $L(x)$ is the illumination determining the rate of reproduction of the bacteria. To solve Eq. (5.13) is a two-step process: First, we find the position $x'(x, t)$ at time $t$ of the fluid element that is at $x$ at time $t = 0$. It satisfies

$$\frac{dx'(x,t)}{dt} = v(x') \tag{5.14}$$

with the initial condition $x'(x, 0) = x$. The solution of the original equation is then

$$\rho(x,t) = f\big(x'(x,-t)\big)\exp\left[\int_{-t}^{0} dt' L\big(x'(x,t)\big)\right] \tag{5.15}$$

where $f(x)$ is the initial density distribution at $t = 0$.

To go from this hydrodynamical–bacteriological analog model to Eq. (5.10) we merely identify $x, t, L, \rho, v \to g, -t, -\gamma, \Gamma$, and $\beta$ and arrive at the solution

$$\Gamma(\lambda p_i, g) = \Gamma(p_i, \bar{g})\exp\left[\int_0^t dt'\gamma\big(\bar{g}(g,t)\big)\right] \tag{5.16}$$

where $\bar{g}$ is the sliding coupling constant, which satisfies

$$\frac{d\bar{g}(g,t)}{dt} = \beta(\bar{g}) \tag{5.17}$$

subject to the boundary condition $\bar{g}(g, 0) = g$.

We may rewrite Eq. (5.16) as

$$\Gamma(\lambda p_i, g) = \Gamma(p_i, \bar{g})\exp\left[\int_g^{\bar{g}} dg' \frac{\gamma(g')}{\beta(g')}\right] \tag{5.18}$$

Provided that $g^2 \ll 1$, we may calculate $\gamma$ and $\beta$ in the lowest-order one-loop of perturbation theory to estimate accurately the exponent in Eq. (5.18). The criterion $\bar{g}^2 \ll 1$ replaces the separate conditions

$$\bar{g}^2 \ll 1 \quad \text{and} \quad g^2 \ln \frac{q^2}{\mu^2} \ll 1 \tag{5.19}$$

necessary in conventional perturbation theory, where large logarithms appear in each order. Thus the renormalization group enables summation of all leading logarithms.

## 5.3
## QCD Asymptotic Freedom

We now calculate the $\beta$-function to one-loop order very explicitly. The principal interest is QCD with gauge group $SU(3)$ and quarks in the **3** representation. The calculation will, however, be done for an arbitrary simple gauge group and arbitrary representations of fermion and scalar fields. We shall also quote from the literature the corresponding results for two- and three-loops.

Consider first a pure Yang–Mills theory without matter fields, only gluons. For this theory we calculate the renormalization constants, $Z_3$ corresponding to the field renormalization of the gluon as well as $Z_1$ for the triple–gluon coupling, respectively. With these conventions the gauge coupling is renormalized according to

$$g_r = \frac{Z_3^{3/2}}{Z_1} g_u \tag{5.20}$$

since each gluon leg carries a factor $Z_3^{1/2}$ at the three-gluon vertex. Thus the $\beta$-function is given by

$$\beta = \mu \frac{\partial g_r}{\partial \mu} = -g_u \frac{\partial}{\partial(\ln \Lambda)} \frac{Z_3^{3/2}}{Z_1} \tag{5.21}$$

since the $Z_i$ depend only on the ratio $\Lambda/\mu$,

Let $\mathscr{Z}_3$, $\mathscr{Z}_1$ denote the renormalization constants for the ghost field and the ghost–gluon vertex, respectively. (Note that $\mathscr{Z}_3$ was denoted simply by $\mathscr{Z}$ in our earlier proof of renormalizability.) Then according to the Taylor–Slavnov identify, one has [29, 30]

$$Z_1 = Z_3 \left( \frac{\mathscr{Z}_1}{\mathscr{Z}_3} \right) \tag{5.22}$$

and this provides a simpler method, using the usual covariant gauges, of calculating $Z_1$. Somewhat later we shall exhibit a simpler calculation of the $\beta$-function in

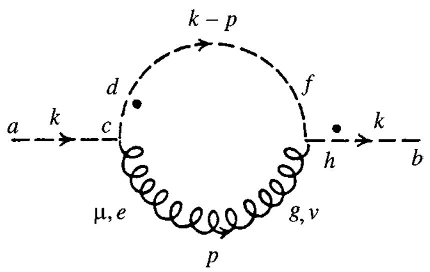



**Figure 5.2** Graph contributing to $\mathscr{Z}_3$.

background field gauge, but here it is useful to calculate these four renormalization constants up to order $g^2$ (one loop) in a general covariant gauge.

Consider first $\mathscr{Z}_3$, for which the only graph is Fig. 5.2. The Feynman rules give for this [here $C_2(G)$ is the quadratic Casimir $C_2(G)\delta_{ab} = \sum_{c,d} C_{acd}C_{bcd}$]

$$\frac{\delta^{ab}C_2(G)}{k^4}\frac{g^2}{16\pi^4}I(k) \tag{5.23}$$

with the integral

$$I(k) = \int \frac{d^4p(k-p)_\mu k_\nu}{(k-p)^2p^2}\left[g_{\mu\nu} - \frac{(1-\alpha)p_\mu p_\nu}{p^2}\right] \tag{5.24}$$

The first term in brackets gives, using Feynman parameters and dimensional regularization,

$$\int_0^1 dx \frac{d^np(k^2 - k\cdot p)}{(p^2 - 2xk\cdot p + xk^2)^2} \simeq i\pi^2k^2 \ln\Lambda \tag{5.25}$$

The second term of Eq. (5.24) becomes

$$-2(1-\alpha)\int_0^1 \frac{dx(1-x)d^np[(k\cdot p)^2 - p^2k\cdot p]}{(p^2 - 2xk\cdot p + xk^2)^3} \simeq i\pi^2(1-\alpha)\frac{k^2}{2}\ln\Lambda \tag{5.26}$$

Thus

$$I(k) \simeq i\pi^2k^2\left(\frac{3}{2} - \frac{\alpha}{2}\right)\ln\Lambda + \text{finite terms} \tag{5.27}$$

and the full diagram becomes

$$\frac{i\delta^{ab}}{k^2}\left(\frac{g}{4\pi}\right)^2\left(\frac{3}{2} - \frac{\alpha}{2}\right)C_2(G)\ln\Lambda + \cdots \tag{5.28}$$

whence

$$\mathscr{Z}_3 = 1\frac{g^2}{16\pi^2}\left(\frac{3}{2} - \frac{\alpha}{2}\right)C_2(G)\ln\Lambda + O\left(g^4\right) \tag{5.29}$$

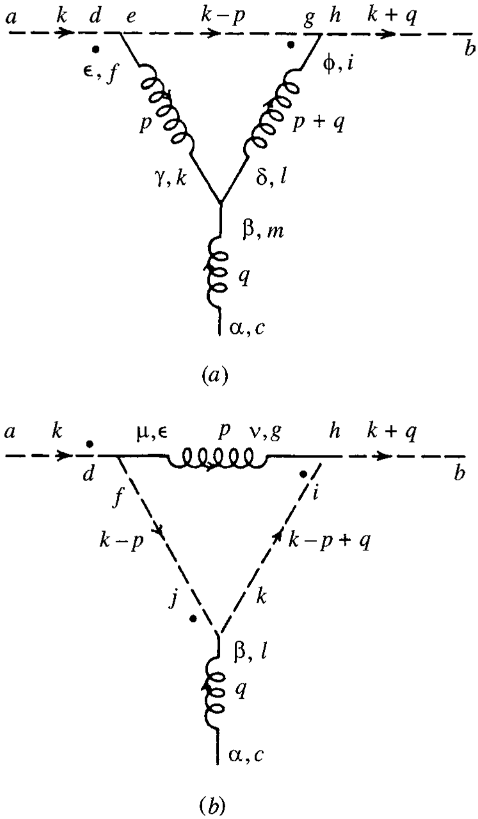



**Figure 5.3** Graphs contributing to $\mathscr{Z}_1$.

The ghost–gluon coupling renormalization constant $\mathscr{Z}_1$ needs evaluation of two Feynman diagrams corresponding to Fig. 5.3*a* and *b*. Figure 5.3*a* gives an amplitude

$$\frac{-ig^3}{16\pi^2} C_{cab} C_2(G) k_\epsilon I_{\epsilon\beta}(k,q) \tag{5.30}$$

where the integral is convergent for the Landau gauge $\alpha = 0$, and hence the divergence is related to the Feynman gauge $\alpha = 1$ by

$$I_{\epsilon\beta}(k,q) \approx \alpha I^F_{\epsilon\beta}(k,q) \tag{5.31}$$

$$I^F_{\epsilon\beta}(k,q) = \int \frac{d^4p}{(k-p)^2 p^2 (p+q)^2} (p-k)_\delta \big[(-2p-q)_\beta g_{\epsilon\delta} + (p+2q)_\epsilon g_{\beta\delta} + (-q+p)_\delta g_{\beta\epsilon}\big] \tag{5.32}$$

$$\approx 2 \int \frac{dx\, dy\, d^4p}{(\text{denominator})^3} p_\delta (-2p_\beta g_{\epsilon\delta} + p_\epsilon g_{\beta\delta} + p_\delta g_{\beta\epsilon}) \tag{5.33}$$

where we keep only the divergent terms; note that the Feynman denominator cancels in the leading logarithmic divergence, so is not needed explicitly. Dimensional regularization then gives

$$I^F_{\epsilon\beta} \sim \frac{3}{2} i\pi^2 g_{\epsilon\beta} \ln \Lambda \tag{5.34}$$

and hence the incremental contribution to $\mathscr{Z}_1$ is

$$\Delta \mathscr{Z}_1 \text{ (Fig. 5.3}a) = \frac{g^2 C_2(G)}{16\pi^2}\left(-\frac{3\alpha}{2}\right) \ln \Lambda \tag{5.35}$$

The second diagram for $\mathscr{Z}_1$, depicted in Fig. 5.3*b*, gives the contribution

$$\frac{ig^3}{16\pi^4} k_\mu \alpha I^F_{\mu\alpha}(k, q) C_2(G) C_{cab} \tag{5.36}$$

where the integral in Feynman gauge $\alpha = 1$ (this diagram also has no logarithmic divergence in Landau gauge $\alpha = 0$) is

$$I^F_{\mu\alpha} = 2 \int dx \frac{dy\, d^n p\, p_\alpha p_\mu}{(\text{denominator})^3} \tag{5.37}$$

$$= \frac{1}{2} i\pi^2 g_{\mu\alpha} \ln \Lambda \tag{5.38}$$

This provides the incremental contribution

$$\Delta \mathscr{Z}_1 \text{ (Fig. 5.3}b) = \frac{g^2 C_2(G)}{16\pi^2}\left(+\frac{\alpha}{2}\right) \ln \Lambda \tag{5.39}$$

and hence we arrive at

$$\mathscr{Z}_1 = 1 + \frac{g^2}{16\pi^2} C_2(G)(-\alpha) \ln \Lambda + O\left(g^4\right) \tag{5.40}$$

Finally, we turn to the gauge field renormalization constant $Z_3$ (recall that $Z_1$ will be deduced from the Taylor–Slavnov identity), for which the relevant Feynman diagrams are shown in Fig. 5.4. Figure 5.4*a* has a quadratic divergence but no logarithmic divergence, hence does not contribute to $Z_3$; in terms of dimensional regularization we state that $\int d^4 p / p^2 = 0$.

Turning to Fig. 5.4*b* we find [bearing in mind the $(-1)$ factor for the ghost loop] the amplitude

$$\delta_{ab} \frac{C_2(G) g^2}{64\pi^4} I_{\alpha\beta}(k) \tag{5.41}$$

where the integral is

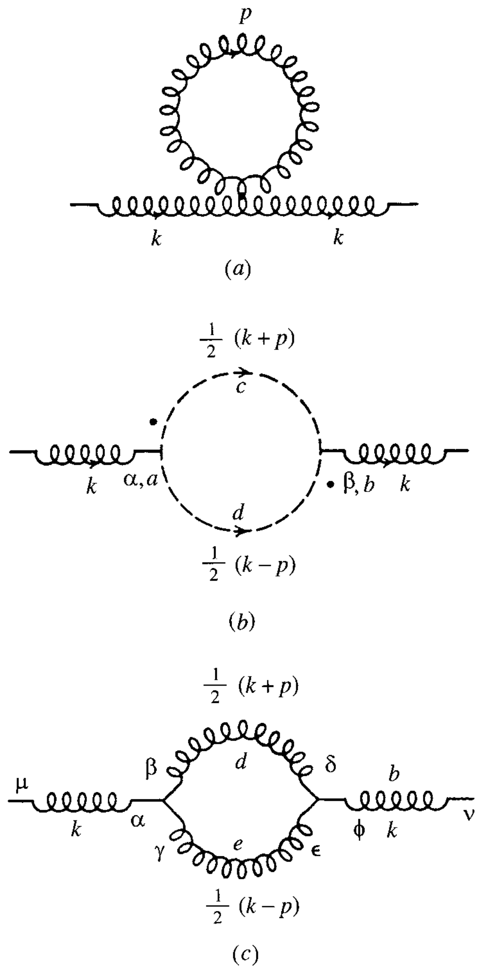



**Figure 5.4** Graphs contributing to $Z_3$.

$$I_{\alpha\beta}(k) = \int \frac{d^4 p(k+p)_\alpha (k-p)_\beta}{(k+p)^2(k-p)^2} \tag{5.42}$$

$$= \int_0^1 dx \frac{d^n p(k+p)_\alpha (k-p)_\beta}{[p^2 + 2p\cdot k(2x-1) + k^2]^2} \tag{5.43}$$

$$\approx i\pi^2 \ln\Lambda \left(\frac{4}{3}k_\alpha k_\beta + \frac{2}{3} g_{\alpha\beta} k^2\right) \tag{5.44}$$

Hence the full diagram gives

$$\frac{-ig^2}{16\pi^2} C_2(G)\delta^{ab} \ln\Lambda \left(\frac{1}{3}k_\alpha k_\beta + \frac{1}{6} g_{\alpha\beta} k^2\right) \tag{5.45}$$

This diagram by itself is not gauge invariant and only becomes so on addition to Fig. 5.4*c*, to which we now turn. Remembering a combinatorial factor of $\frac{1}{2}$ for the identical bosons, the amplitude is

$$\delta_{ab}\frac{g^2}{16\pi^4}C_2(G)I_{\alpha\phi}(k) \tag{5.46}$$

with the integral

$$\begin{aligned} I_{\alpha\phi}(k) &= \frac{1}{2}\int \frac{d^4p}{(k+p)^2(k-p)^2}\Pi_{\beta\delta}(k+p)\Pi_{\gamma\epsilon}(k-p) \\ &\cdot \Gamma_{\alpha\beta\gamma}\left(k, \frac{-k-p}{2}, \frac{-k+p}{2}\right)\Gamma_{\phi\delta\epsilon}\left(-k, \frac{k+p}{2}, \frac{k-p}{2}\right) \end{aligned} \tag{5.47}$$

in which we defined

$$\Pi_{\mu\nu}(k) = g_{\mu\nu} - \frac{(1-\alpha)k_\mu k_\nu}{k^2} \tag{5.48}$$

$$\Gamma_{\alpha\beta\gamma}(p,q,r) = g_{\alpha\beta}(q-p)_\gamma + g_{\beta\gamma}(r-q)_\alpha + g_{\gamma\alpha}(p-r)_\beta \tag{5.49}$$

It is convenient to write

$$I_{\alpha\phi}(k) = I^F_{\alpha\phi}(k) + (1-\alpha)I^{(1)}_{\alpha\phi}(k) + (1-\alpha)^2 I^{(2)}_{\alpha\phi}(k) \tag{5.50}$$

where the superscript $F$ means Feynman gauge $\alpha = 1$.

The first remark is that $I^{(2)}$ has no divergence; to see this, note that

$$\Gamma_{\alpha\beta\gamma}\left(k, \frac{-k-p}{2}, \frac{-k+p}{2}\right)(k+p)_\beta(k-p)_\gamma = -2p_\mu\left(k^2 g_{\mu\alpha} - k_\mu k_\alpha\right) \tag{5.51}$$

$$\Gamma_{\phi\delta\epsilon}\left(-k, \frac{k+p}{2}, \frac{k-p}{2}\right)(k+p)_\delta(k-p)_\epsilon = 2p_\mu\left(k^2 g_{\mu\phi} - k_\mu k_\phi\right) \tag{5.52}$$

and hence

$$I^{(2)}_{\alpha\phi}(k) = -2\left(k^2 g_{\mu\alpha} - k_\mu k_\alpha\right)\left(k^2 g_{\nu\phi} - k_\nu k_\phi\right)\int \frac{d^4p\, p_\mu p_\nu}{(k+p)^4(k-p)^4} \tag{5.53}$$

which is ultraviolet finite and hence does not contribute to $Z_3$.

Next we observe that there are two equal contributions to $I^{(1)}_{\alpha\phi}(k)$, coming from the $(1-\alpha)$ terms of the two vector propagators, respectively. Indeed, we may exhibit this equality (which holds only for the symmetric choice of momenta),

$$\int \frac{d^4 p}{(k+p)^4 (k-p)^2} \Gamma_{\alpha\beta\gamma}\left(k, \frac{-k-p}{2}, \frac{-k+p}{2}\right)$$
$$\cdot \Gamma_{\phi\delta\epsilon}\left(-k, \frac{k+p}{2}, \frac{k-p}{2}\right)(k+p)_\beta (k+p)_\delta g_{\gamma\epsilon}$$
$$= \int \frac{d^4 p}{(k+p)^2 (k-p)^4} \Gamma_{\alpha\beta\gamma}\left(k, \frac{-k-p}{2}, \frac{-k+p}{2}\right)$$
$$\cdot \Gamma_{\phi\delta\epsilon}\left(-k, \frac{k+p}{2}, \frac{k-p}{2}\right)(k-p)_\gamma (k-p)_\delta g_{\beta\delta} \tag{5.54}$$

by using the symmetry

$$\Gamma_{\alpha\beta\gamma}(p, q, r) = -\Gamma_{\alpha\gamma\beta}(p, q, r) \tag{5.55}$$

and the fact that the integrals satisfy

$$I_{\alpha\phi}(k) = I_{\alpha\phi}(-k) = I_{\phi\alpha}(k) = I_{\phi\alpha}(-k) \tag{5.56}$$

since by Lorentz invariance there must always be a decomposition

$$I_{\alpha\phi}(k) = f_1(k^2) g_{\alpha\phi} + f_2(k^2) k_\alpha k_\phi \tag{5.57}$$

Thus to compute $I^{(1)}_{\alpha\phi}(k)$ we need compute only, say, the left-hand side of Eq. (5.54), then multiply by 2. Now use for the tensor in the integrand

$$T_{\alpha\phi}(k, p) = \Gamma_{\alpha\beta\gamma}\left(k, \frac{-k-p}{2}, \frac{-k+p}{2}\right)$$
$$\cdot \Gamma_{\phi\delta\epsilon}\left(-k, \frac{k+p}{2}, \frac{k-p}{2}\right)(k+p)_\beta (k+p)_\delta g_{\gamma\epsilon} \tag{5.58}$$

and then Feynman parametrization with dimensional regularization gives

$$I^{(1)}_{\alpha\phi}(k) = 2\int_0^1 \frac{dx\, x d^n p}{[p^2 + 2p\cdot k(2x-1) + k^2]^3} T_{\alpha\phi}(k, p) \tag{5.59}$$

$$\approx -i\pi^2 \ln\Lambda (k_\alpha k_\phi - k^2 g_{\alpha\phi}) \tag{5.60}$$

There remains only the evaluation in Feynman gauge, which needs

$$t^F_{\alpha\phi}(k, p) = \Gamma_{\alpha\beta\gamma}\left(k, \frac{-k-p}{2}, \frac{-k+p}{2}\right)\Gamma_{\phi\delta\epsilon}\left(-k, \frac{k+p}{2}, \frac{k-p}{2}\right) g_{\beta\delta} g_{\gamma\epsilon} \tag{5.61}$$

$$= g_{\alpha\phi}\left(-\frac{9}{2}k^2 - \frac{1}{2}p^2\right) - \frac{5}{2}p_\alpha p_\phi + \frac{9}{2}k_\alpha k_\phi \tag{5.62}$$

whence

$$I^F_{\alpha\phi}(k) = \frac{1}{2}\int_0^1 \frac{dx\, d^n p}{[p^2 + 2p\cdot k(2x-1) + k^2]^2} T^T_{\alpha\phi}(k, p) \tag{5.63}$$

$$\approx -i\pi^2 \ln\Lambda\left(\frac{11}{3}k_\alpha k_\phi - \frac{19}{6}k^2 g_{\alpha\phi}\right) \tag{5.64}$$

The fact that Eq. (5.64) is not gauge covariant is no surprise: It was known in advance from Eq. (5.45) for the ghost loop. Combining the logarithmic divergences from Eqs. (5.45), (5.60), and (5.64) now enables us finally to compute

$$Z_3 = 1 + \frac{g^2}{16\pi^2}C_2(G)\ln\Lambda\left(\frac{13}{3} - \alpha\right) + O\left(g^4\right) \tag{5.65}$$

Together with the expressions for $\mathscr{Z}_1$ and $\mathscr{Z}_3$ in Eqs. (5.40) and (5.29), we may now use the identity of Eq. (5.22) to derive

$$Z_1 = 1 + \frac{g^2}{16\pi^2}C_2(G)\ln\Lambda\left[\left(\frac{13}{3} - \alpha\right) + (-\alpha) - \left(\frac{3}{2} - \frac{\alpha}{2}\right)\right] \tag{5.66}$$

$$= 1 + \frac{g^2}{16\pi^2}C_2(G)\ln\Lambda\left(\frac{17}{6} - \frac{3\alpha}{2}\right) + O\left(g^4\right) \tag{5.67}$$

Direct calculation of the graphs for $Z_1$ depicted in Fig. 5.5 gives the same answer.

For pure Yang–Mills theory, the $\beta$-function follows from Eq. (5.21) as

$$\beta = -g\frac{\partial}{\partial(\ln\Lambda)}\frac{Z_3^{3/2}}{Z_1} \tag{5.68}$$

$$= -g\left(\frac{g^2}{16\pi^2}\right)\frac{11}{3}C_2(G) + O\left(g^5\right) \tag{5.69}$$

The negative sign is the crucial signal for asymptotic freedom—the origin $g = 0$ is now an ultraviolet fixed point of the renormalization group—and, as we shall see, leads to an explanation of the weak coupling between quarks observed at high energies. The result of this calculation was known to 't Hooft in 1972 [12].

We now add fermion (spin-$\frac{1}{2}$) and scalar matter fields, which will contribute positively to $\beta$, and hence the question is: How much matter may we include before the origin $g = 0$ switches from ultraviolet to infrared fixed point and asymptotic freedom is lost?

First consider the spin-$\frac{1}{2}$ loop of Fig. 5.6, which gives

$$T(R)\delta^{ab}\frac{g^2}{16\pi^4}I_{\alpha\beta}(k) \tag{5.70}$$

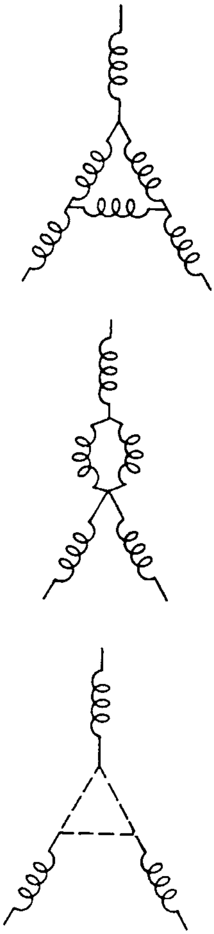


**Figure 5.5** Graphs contributing to $Z_1$.

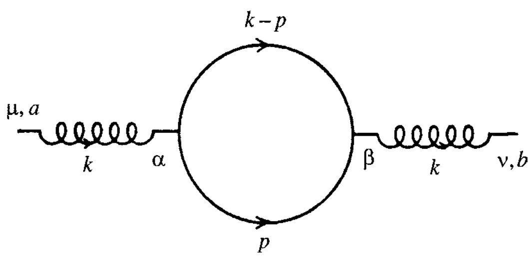



**Figure 5.6** Fermion loop diagram.

with

$$I_{\alpha\beta}(k) = \int \frac{d^4 p}{(k-p)^2 p^2} \, \mathrm{Tr}(\gamma_\alpha \not{k} \gamma_\beta \not{p} - \gamma_\alpha \not{p} \gamma_\beta \not{p}) \tag{5.71}$$

and the index $T(R)$ for an arbitrary representation $R$ of the gauge group $G$ is given by

$$\mathrm{Tr}\big(T^a(R)T^b(R)\big) = \delta^{ab}T(R) \tag{5.72}$$

Here the $T^a(R)$ are generators of $G$ written in the $d(R) \times d(R)$ basis, with $d(R)$ the dimensionality of $R$.

The integral is

$$I_{\alpha\beta}(k) = 4\int_0^1 dx \frac{d^n p}{p^2 - 2xk\cdot p + xk^2} \cdot \big(k_\alpha p_\beta + k_\beta p_\alpha - g_{\alpha\beta}k\cdot p - 2p_\alpha p_\beta + p^2 g_{\alpha\beta}\big) \tag{5.73}$$

$$\approx \frac{8i\pi^2}{3}\ln\Lambda\big(k_\alpha k_\beta - k^2 g_{\alpha\beta}\big) \tag{5.74}$$

Combining this with Eq. (5.70), we find a contribution for Dirac four-component fermions in representation $R$:

$$\Delta Z_1 = \Delta Z_3 = -\frac{g^2}{16\pi^2}\left(\frac{8}{3}\right)T(R)\ln\Lambda \tag{5.75}$$

The fact that $\Delta Z_1 = \Delta Z_3$ (since $\Delta\mathscr{Z}_1 = \Delta\mathscr{Z}_3 = 0$) is an example of a Ward–Takahashi identity [31, 32], as in QED.

For scalar particles, the diagrams are shown in Fig. 5.7. The amplitude for Fig. 5.7*a* is quadratically divergent and gives no contribution to $Z_3$. Figure 5.7*b* gives

$$-g^2\delta^{ab}T(R)I_{\alpha\beta}(k) \tag{5.76}$$

with

$$I_{\alpha\beta}(k) = \int_0^1 dx \frac{d^n p}{(p^2 - 2k\cdot px + xk^2)^2}(k-2p)_\alpha(k-2p)_\beta \tag{5.77}$$

$$\approx \frac{2}{3}i\pi^2(k_\alpha k_\beta) - k^2 g_{\alpha\beta}\ln\Lambda \tag{5.78}$$

Hence

$$\Delta Z_1 = \Delta Z_3 = -\frac{g^2}{16\pi^2}\left(\frac{2}{3}\right)T(R)\ln\Lambda \tag{5.79}$$

as the contribution from complex scalars in representation $R$.

Thus we have

$$Z_1 = 1 + \frac{g^2}{16\pi^2}\ln\Lambda\left[C_2(G)\left(\frac{17}{6} - \frac{3\alpha}{2}\right) - \frac{8}{3}T(R_D) - \frac{2}{3}T(R_C)\right] \tag{5.80}$$

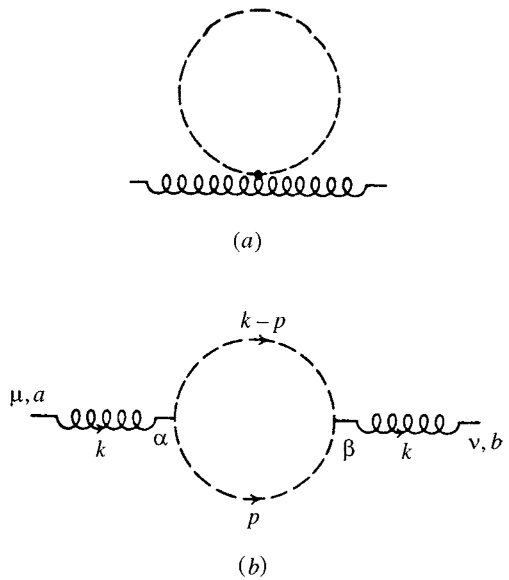



Figure 5.7 Scalar loop diagrams.

$$Z_3 = 1 + \frac{g^2}{16\pi^2} \ln \Lambda \left[ C_2(G) \left( \frac{13}{3} - \alpha \right) - \frac{8}{3} T(R_D) - \frac{2}{3} T(R_C) \right] \tag{5.81}$$

and hence

$$\beta = -g \frac{\partial}{\partial (\ln \Lambda)} \frac{Z_3^{3/2}}{Z_1} \tag{5.82}$$

$$= -\frac{g^3}{16\pi^2} \left[ \frac{11}{3} C_2(G) - \frac{4}{3} T(R_D) - \frac{1}{3} T(R_C) \right] \tag{5.83}$$

where the subscripts $D$ and $C$ refer to Dirac fermions and complex scalars, respectively. For Weyl fermions or for real scalars, the contribution would be halved; that is,

$$\beta = -\frac{g^3}{16\pi^2} \left[ \frac{11}{3} C_2(G) - \frac{2}{3} T(R_W) - \frac{1}{6} T(R_R) \right] \tag{5.84}$$

where the subscripts $W$ and $R$ refer to Weyl and real, respectively.

Note that although $Z_1$ and $Z_3$ are gauge dependent, $\beta$ is not; this is true to all orders of perturbation theory [33, 34]. Also note that for any irreducible representation, we may identify the index as

$$T(R) = \frac{C_2(R) d(R)}{r} \tag{5.85}$$

where $C_2(R)$ is the quadratic Casimir, $d(R)$ is the dimension, and $r$ is the dimension of the group [$r = N^2 - 1$ for SU($N$)].

Let us first consider QCD for which $G = \mathrm{SU}(3)$, $C_2(G) = 3$, and $T(R_M) = N_f$, the number of quark flavors. In lowest order,

$$\beta = -\frac{g^3}{16\pi^2}\left(11 - \frac{2}{3}N_f\right) \tag{5.86}$$

which is negative for $N_f \leqslant 16$, showing that we may add up to 16 quark flavors without losing asymptotic freedom. The calculation has been extended to two loops [35, 36] and to three-loops [37–39], with the result [39]

$$\begin{aligned}\beta = &-\frac{g^3}{16\pi^2}\left(11 - \frac{2}{3}N_f\right) - \frac{g^5}{(16\pi^2)^2}\left(102 - \frac{38}{3}N_f\right)\\ &- \frac{g^7}{(16\pi^2)^3}\left(\frac{2857}{2} - \frac{5033}{18}N_f + \frac{325}{54}N_f^2\right) + O\left(g^9\right)\end{aligned} \tag{5.87}$$

Another interesting case is supersymmetric Yang–Mills theory, with a chiral matter superfield in the adjoint representation. From Eq. (5.84) we then have at lowest order

$$\beta = -\frac{g^3}{16\pi^2}\frac{C_2(G)}{6}\left[22 - 4\nu(M) - \nu(R)\right] \tag{5.88}$$

where $\nu(M)$ and $\nu(R)$ are the numbers of Majorana fermions and real scalars, respectively. The gauge supermultiplet contains (for unextended $N = 1$ supersymmetry) one Majorana fermion, while each chiral superfield contains one Majorana fermion and two real scalars. Thus for $N = 1$ and $n$ chiral superfields,

$$\beta = -\frac{g^3}{16\pi^2}\frac{C_2(G)}{6}(18 - 6n) \tag{5.89}$$

which vanishes for $n = 3$. This was noticed by Ferrara and Zumino [40] and by Salam and Strathdee [41], who all wondered about higher loops. The two-loop result for $N - 1$ supersymmetry was computed by Jones [42], with the result

$$\beta = -\frac{-g^3}{16\pi^2}\frac{C_2(G)}{6}(18 - 6n) - \frac{g^5}{(16\pi^2)^2}C_2(G)^2(6 - 10n + 4n) \tag{5.90}$$

which fails to vanish for $n = 3$. In the final set of parentheses, the $-10n$ is from gauge couplings and the $+4n$ is from Yukawa couplings for vanishing superpotential. Hence this fails to generalize. However, a more geometrical theory is $\mathcal{N} = 4$ extended supersymmetric Yang–Mills, where in Eq. (5.88) one has $\nu(M) = 4$ and $\nu(R) = 6$, so that $\beta$ vanishes at one loop. In two-loops, the Yukawa couplings give $8n$ rather than $4n$ with $n = 3$ in Eq. (5.90), and hence the two-loop $\beta$-function vanishes, too [43, 44]. In fact, at three-loops [45, 46], the $\beta$-function remains zero. It has even been shown [47] that $\mathcal{N} = 4$ extended supersymmetric Yang–Mills theory is finite to all orders of perturbation theory. In $\mathcal{N} = 2$ extended supersymmetric Yang–Mills, if the one-loop $\beta$-function vanishes the theory is finite to all orders [48].

For the $\mathcal{N} = 1$ case (the most phenomenologically interesting because it allows chiral fermions) vanishing of the one-loop $\beta$-function implies two-loop finiteness [49, 50]; for this $\mathcal{N} = 1$ case, however, the finiteness does not persist, in general, at the three-loop level [51, 52].

## 5.4 Grand Unification

As already mentioned, the first triumph for the renormalization group applied to particle theory was the explanation of approximate scaling using QCD. Chronologically, second application involved grand unification [20].

Consider the low-energy ($\lesssim$ 100 GeV) phenomenological gauge group

$$\mathrm{SU}(3) \times \left[\mathrm{SU}(2) \times \mathrm{U}(1)\right] \tag{5.91}$$

with gauge couplings $g_i(\mu)$, $i = 3, 2, 1$, in an obvious notation. We may take the phenomenological values at some $\mu \sim 100$ GeV, say, then study how they evolve according to the renormalization group, which gives at lowest one-loop order

$$\mu \frac{d}{d\mu} g_i(\mu) = b_i g_i^3(\mu) \tag{5.92}$$

with, for $\mathrm{SU}(N)$,

$$b = -\frac{1}{16\pi^2}\left[\frac{11}{3}N - \frac{2}{3}T(R_W)\right] \tag{5.93}$$

where we include Weyl fermions but ignore scalars. Thus we have, for $\mathrm{SU}(N)$, $T(R_M = \mathbf{N}) = \frac{1}{2}$. Each quark–lepton family has four triplets of color $\mathrm{SU}(3)$ and four doublets of electro weak $\mathrm{SU}(2)$, giving

$$b_3 = -\frac{1}{16\pi^2}\left(11 - \frac{2}{3}N_f\right) \tag{5.94}$$

$$b_2 = -\frac{1}{16\pi^2}\left(\frac{22}{3} - \frac{2}{3}N_f\right) \tag{5.95}$$

Since in a grand unifying group, $G$, the contribution of a complete irreducible fermion representation of $G$ to the vacuum polarization will be the same for any gauge particle of $G$, the equality visible in Eqs. (5.94) and (5.95) is already suggestive of

$$\mathrm{SU}(3) \times \left[\mathrm{SU}(2) \times \mathrm{U}(1)\right] \subset G \tag{5.96}$$

such that each quark–lepton family forms one or more complete irreducible representations of $G$.

Suppose that we define weak hypercharge $Y$ by $Q = T_3 + Y$; then $T(R_M)$ for the corresponding $U(1)_Y$ will be given by the eigenvalue $Y^2$. In fact, the normalization of the U(1) generator is more arbitrary, so we write $Q = T_3 + CY$. Each family has six states with $Y = \frac{1}{6}$, three each with $Y = -\frac{2}{3}$ and $+\frac{1}{3}$, two with $Y = -\frac{1}{2}$, and one with $Y = +1$. Thus

$$b_1 = \frac{1}{16\pi^2} \cdot \frac{2}{3} N_f \frac{5}{3C^2} \tag{5.97}$$

This suggests $C^2 = \frac{5}{3}$ as the appropriate identification for grand unification. Indeed, taking $G = SU(5)$ and each family as $(10 + \bar{5})_L$, it is immediate to check that since $\mathrm{Tr}(T_3^2) = \mathrm{Tr}(Y^2) = \frac{1}{2}$ and $\mathrm{Tr}(T_3 Y) = 0$, then

$$\mathrm{Tr}\big(Q^2\big) = \mathrm{Tr}(T_3 + CY)^2 = \frac{1}{2}\big(1 + C^2\big) = \frac{4}{3} \tag{5.98}$$

since in the defining representation $Q = \begin{pmatrix} -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & +1 \end{pmatrix}$, corresponding to $5 = \begin{pmatrix} d_1 & d_2 & d_3 & \bar{\nu}_e & e^+ \end{pmatrix}$, for example. Thus we see that

$$b_2 = b_1 - \frac{11}{24\pi^2} \tag{5.99}$$

$$b_3 = b_1 - \frac{11}{16\pi^2} \tag{5.100}$$

At the unification mass $M$ the three couplings $g_i(\mu)$ ($i = 1, 2, 3$) become equal:

$$g_1(M) = g_2(M) = g_3(M) = g_G(M) \tag{5.101}$$

Thus, the unification enables us to determine the electroweak mixing angle $\sin^2\theta_W$ evaluated at mass $M$. In the notation of Weinberg [53], who denotes the gauge couplings of $SU(2)_L$ and $U(1)_Y$ as $g$ and $g'$, respectively, the electroweak mixing angle is

$$\sin^2\theta_W(\mu) = \frac{g'^2(\mu)}{g^2(\mu) + g'^2(\mu)} \tag{5.102}$$

Here we have acknowledged the fact that $\theta_W(\mu)$ is dependent on the energy scale by virtue of the renormalization group. We now observe that $g = g_2$ while $g' = g_1/C$, and hence that

$$\sin^2\theta_W(M) = \frac{1}{1 + C^2} = \frac{1}{2\,\mathrm{Tr}(Q^2)} \tag{5.103}$$

and, in particular, $\sin^2\theta_W(M) = \frac{3}{8}$[21] for SU(5), where $C^2 = \frac{5}{3}$.

Provided that the $g_i(\mu)$ remain sufficiently small so that perturbation theory is reliable, we may solve Eq. (5.92) to give, for $\mu \leqslant M$ (we follow Ref. [20] precisely here),

$$g_i^{-2}(\mu) = g_G^{-2}(M) + 2b_i \ln \frac{M}{\mu} \tag{5.104}$$

Assuming that only complete representations of $G$ make up the light fermions (masses $m \lesssim 100$ GeV) and that there are no fermions with masses in the "desert" energy region 100 GeV $\lesssim m \leqslant M$, we may use Eqs. (5.99) and (5.100) for the entire interpolation between 100 GeV and $M$. Defining the sliding electric charge of the electron by

$$e^2(\mu) = g_2^2(\mu) \sin^2 \theta_W(\mu) \tag{5.105}$$

we find by straightforward algebra that

$$\frac{C^2}{g_1^2(\mu)} + \frac{1}{g_2^2(\mu)} - \frac{1+C^2}{g_3^2(\mu)} = \frac{1}{e^2(\mu)} - \frac{1+C^2}{g_3^2(\mu)} \tag{5.106}$$

$$= 2 \ln \frac{M}{\mu} \left[ b_1 C^2 + b_2 - b_3 \left(1 + C^2\right) \right] \tag{5.107}$$

and

$$C^2 \left[ g_1(\mu)^{-2} - g_2(\mu)^{-2} \right] = e(\mu)^{-2} \left[ 1 - \left(1 + C^2\right) \sin^2 \theta_W(\mu) \right] \tag{5.108}$$

$$= 2C^2 (b_1 - b_2) \ln \frac{M}{\mu} \tag{5.109}$$

Substitution of $(b_1 - b_2)$ and $(b_1 - b_3)$ from Eqs. (5.99) and (5.100) then gives the final formulas for $M$ and $\sin^2 \theta_W(\mu)$ as follows:

$$\ln \frac{M}{\mu} = \frac{24\pi^2}{11(1+3C^2)} \left[ \frac{1}{e^2(\mu)} - \frac{1+C^2}{g_3(\mu)^2} \right] \tag{5.110}$$

$$\sin^2 \theta_W(\mu) = \frac{1}{(1+3C^2)} \left[ 1 + \frac{2C^2 e^2(\mu)}{g_3(\mu)^2} \right] \tag{5.111}$$

For $g_3(\mu)$ consider the one-loop QCD formula

$$\alpha_s(\mu) = \frac{g_3(\mu^2)}{4\pi} = \frac{12\pi}{(33 - 2N_f) \ln(\mu^2/\Lambda^2)} \tag{5.112}$$

The value of $\Lambda$, the QCD scale, needs to be determined phenomenologically. Because of precocious scaling, we know that for $\mu^2 = 2$ GeV$^2$ the parameter $\alpha_s/\pi$ which governs the convergence of perturbation theory must be small ($\leqslant 0.3$); on the other hand, perturbation theory must break down at the charge radius of, say, the proton $r_p = 0.8f$ or $\mu^2 = 0.3$ GeV, so here $\alpha_s/\pi \geqslant 0.3$. Putting $N_f = 4$, this gives the allowed range

$$200 \text{ MeV} \lesssim \Lambda \lesssim 600 \text{ MeV} \tag{5.113}$$

Analyses favor a value $\Lambda \sim 200$ MeV (see, e.g. Ref. [54]). Let us take $\mu = 100$ GeV as our low-energy scale, sufficiently close to the mass of the weak gauge bosons $W^{\pm}$ and $Z^0$. Then for $\Lambda = 200$ MeV, $\alpha_s(\mu) = 0.12$, and $g_3(\mu)^2 = 1.51$.

Next consider $e(\mu)^2$. In Ref. [20], Georgi, Quinn, and Weinberg used simply $e(\mu)^2/4\pi = \alpha^{-1} = (137.036)^{-1}$, but as pointed out in Refs. [55] and [56], this value is too low since it also runs with increasing energy. The appropriate formula is, for $\alpha_e(\mu) = e(\mu)^2/4\pi$,

$$\alpha^{-1}(\mu) = \alpha^{-1} - \frac{2}{3\pi}\sum_f Q_f^2 \ln\frac{\mu}{m_f} + \frac{1}{6\pi} \tag{5.114}$$

where the sum is over all fermions, leptons, and quarks with $m_f < \mu$ (a color factor of 3 is necessary for quarks) and $Q_f$ is the fermion's electric charge. For $\mu = 100$ GeV this gives $\alpha^{-1}(\mu) \simeq 128$ (see also Rev. [57]), $e(\mu)^2 = 0.10$.

We are now ready to substitute into Eqs. (5.110) and (5.111) with the results, for SU(5), that

$$M = 7 \times 10^{14} \text{ GeV} \tag{5.115}$$

$$\sin^2\theta_W = 0.203 \tag{5.116}$$

In SU(5) [21] there are superheavy-gauge bosons with mass $M$ that mediate proton decay. By dimensional considerations we may estimate

$$\tau_p = \frac{1}{\alpha_G(M)^2}\frac{M^4}{M_p^5} = 1.1 \times 10^{32} \text{ years} \tag{5.117}$$

with a large uncertainty (at least an order of magnitude) due to the long extrapolation in energy and the occurrence of $M$ to the fourth power. The values in Eqs. (5.116) and (5.117) are now excluded by experiment.

## 5.5 Scaling Violations

Returning now to QCD, the significance of asymptotic freedom is that the origin in coupling constant space $g = 0$ becomes an ultra violet fixed point of the renormalization group. Consider $\beta$ as a function of $g$ and the possible behaviors exhibited in Fig. 5.8. Bearing in mind that $\beta = \mu\, dg/d\mu$, we see that in Fig. 5.8*a* the value $g_1$ is approached by $g$ for increasing $\mu$ provided that we start from $0 < g < g_2$. Similarly, $g_3$ is approached from $g_2 < g < g_4$ and $g_5$ from $g_4 < g < \infty$. Thus

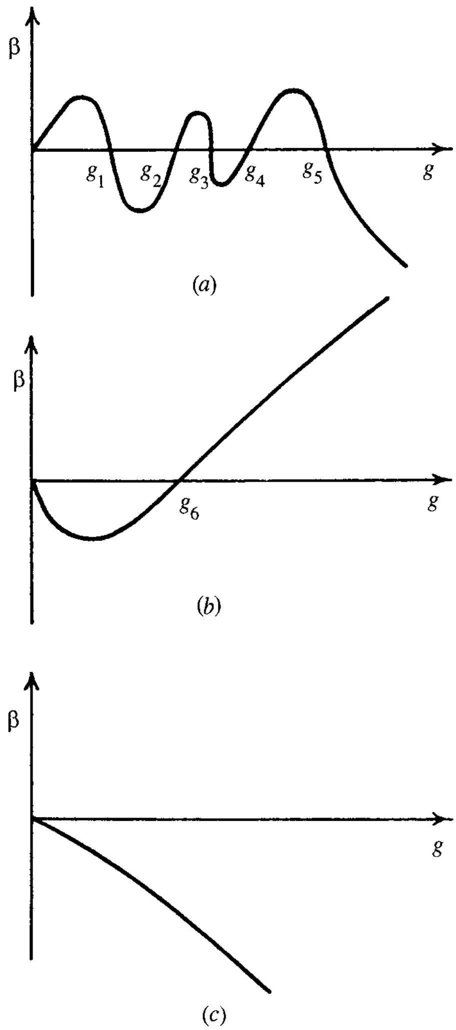



**Figure 5.8** Some possible behaviors of $\beta(g)$.

$g_1$, $g_3$, and $g_5$ in Fig. 5.8*a* are ultraviolet fixed point. Conversely, $g = 0$, $g_2$, and $g_4$ in Fig. 5.8*a* are infrared fixed points approached by $g(\mu)$ for decreasing $\mu$. The general picture is clear: If the slope $\beta'(g)$ is positive at $\beta = 0$, there is an infrared fixed point; if $\beta'$ is negative at $\beta = 0$, there is an ultraviolet fixed point.

In Fig. 5.8*b* the origin is an ultraviolet fixed point, and any $g$ such that $0 < g < g_6$ decreases toward zero at high energy (i.e., asymptotic freedom). At low energies, $g$ increases to $g_6$.

The situation of QCD is depicted in Fig. 5.8*c*, where the origin is asymptotically free (as we have proved above for 16 flavors) and at low energy the coupling eventually becomes infinite. Such low-energy behavior has not been proven but is expected since it yields color confinement, and strongly supported by lattice calculations.

What is so physically compelling about the asymptotic freedom of QCD? We have derived in Eq. (5.16) that Green's functions transform under scaling $p_i \to \lambda p_i$ according to

$$\Gamma(\lambda p, g) = \Gamma(p, \bar{g}) \exp\left[\int_0^t \gamma\big(\bar{g}(g, t')\big)\, dt'\right] \tag{5.118}$$

Keeping at lowest order $\beta = -bg^3$ and $\gamma = cg^2$, we then see that

$$\bar{g}^2(t) = \frac{g^2}{1 + 2bg^2 t} \tag{5.119}$$

where $g = \bar{g}(0)$ and hence the exponential factor in Eq. (5.118) becomes

$$\exp\left(cg^2 \int_0^t \frac{dt'}{1 + 2bg^2 t'}\right) = \left(\frac{g^2}{\bar{g}^2}\right)^{c/2b} \tag{5.120}$$

$$\approx (\ln \lambda)^{c/2b} \tag{5.121}$$

so that we have logarithmic corrections to scaling. Because of this, high-energy processes give a unique testing ground for perturbative QCD.

The processes that have been checked against QCD predictions include:

1. Electron–positron annihilation, particularly the total cross section and production of jets.
2. Deep-inelastic lepton–hadron scattering, particularly scaling violations.
3. Hadron–hadron scattering, particularly large transverse momentum processes and lepton pair production.
4. Decays of heavy quarkonium systems.

We shall presently concentrate on inclusive lepton–hadron scattering,

$$l(p_l) + N(p) \to l'(p_{l'}) + X(p_x) \tag{5.122}$$

since it will sufficiently illustrate use of both the $\beta$ and $\gamma$ (anomalous dimension) functions. Following tradition, we define $q = p_{l'} - p_l$ and $Q^2 = -q^2$ as the "spacelike squared mass" of the exchanged vector boson, be it any of the four electroweak intermediaries $\gamma$, $W^{\pm}$, or $Z^0$. We also define $\nu = (p \cdot q)/M_N$. In the laboratory frame where the target nucleon is at rest, $\nu = E' - E$ is the energy transfer.

As indicated in Fig. 5.9, the squared matrix element is written $L_{\mu\nu} W_{\mu\nu}$, where $L_{\mu\nu}$ refers to the well-established leptonic part while $W_{\mu\nu}$ is expressed in terms of three structure functions $W_i(\nu, Q^2)$ as follows:

$$W_{\mu\nu} = \frac{1}{4\pi} \int d^4z\, e^{iq\cdot z} \langle p|\big[J_\mu^\dagger(z), J_\nu(0)\big]|p\rangle \tag{5.123}$$

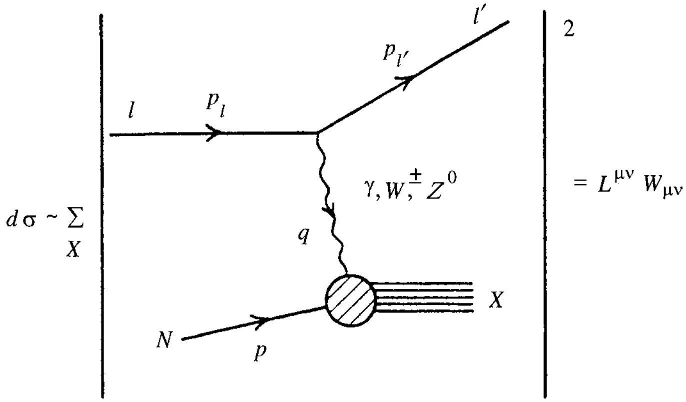



**Figure 5.9** Squared matrix element for lepton–nucleon scattering.

$$= \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right)_1 + \frac{1}{M_N^2}\left(p_\mu - \frac{p\cdot q}{q^2}q_\mu\right)\left(p_\nu - \frac{p\cdot q}{q^2}q_\nu\right)W_2$$

$$- i\frac{\epsilon_{\mu\nu\alpha\beta}p^\alpha q^\beta}{2M_N^2}W_3 \tag{5.124}$$

The third structure function, $W_3$, contributes only to parity violating effects such as A–V interference and hence is absent for photon exchange. Indeed, for deep-inelastic electroproduction, the differential cross section is given by

$$\frac{d\sigma}{dQ^2 d\nu} = \frac{4\pi\alpha^2}{M_N Q^4}\frac{E'}{E}\left(2W_1\sin^2\frac{\theta}{2} + W_2\cos^2\frac{\theta}{2}\right) \tag{5.125}$$

For charged $W^\pm$ exchange (mass $M_W$) the corresponding formulas for $\nu(\bar\nu) + N \to \mu^-(\mu^+) + X$ are

$$\frac{d\sigma^{\nu(\bar\nu)}}{dQ^2 d\nu} = \frac{G_F^2}{2\pi M_N}\frac{E'}{E}\left(\frac{M_W^2}{Q^2 + M_W^2}\right)^2$$

$$\cdot\left(2W_1\sin^2\frac{\theta}{2} + W_2\cos^2\frac{\theta}{2} \mp \frac{E+E'}{M}W_3\sin^2\frac{\theta}{2}\right) \tag{5.126}$$

In Eqs. (5.125) and (5.126), $\theta$ is the scattering angle of the lepton given by $Q^2 = 4EE'\sin^2(\theta/2)$.

Further kinematic variables that may be useful are $x = \omega^{-1} = Q^2/2M_N\nu$, $y = \nu/E$, and $W^2 = (p+q)^2 = p_x^2$. Since $W^2 \geqslant M_N^2$, we have that $0 \leqslant x \leqslant 1$. Also, $y$ is the fraction of the energy transferred to the hadronic system in the laboratory frame. The deep inelastic regime is defined as that where $Q^2$, $M_N\nu$, $W^2 \gg M_N^2$ at fixed $x$, and the scaling phenomenon, sometimes termed Bjorken scaling [59–61], is the experimental observation that in this limit

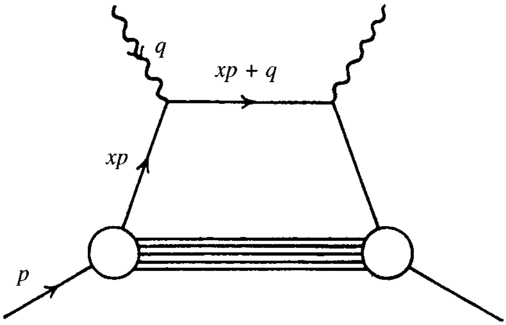



**Figure 5.10** Handbag diagram.

$$W_1(\nu, Q^2) \to F_1(x) \tag{5.127}$$

$$\frac{\nu}{M_N} W_{2,3}(\nu, Q^2) \to F_{2,3}(x) \tag{5.128}$$

so that the structure functions depend only on the dimensionless ratio $x$ (scaling variable) of the two available variables. The naive quark–parton model is defined in this deep inelastic region and is based on the impulse approximation. (For an excellent introductory review, see, e.g., Ref. [62].) The constituent quarks define an electromagnetic source

$$J_\mu^{\text{em}} = \frac{2}{3}\bar{u}\gamma_\mu u - \frac{1}{3}\bar{d}\gamma_\mu d - \frac{1}{3}\bar{s}\gamma_\mu s + \cdots \tag{5.129}$$

and a charged weak source (neglecting charm and higher flavors)

$$J_\mu^{\text{weak}} = \cos\theta_c \bar{u}\gamma_\mu(1-\gamma_5)d + \sin\theta_c \bar{u}\gamma_\mu(1-\gamma_5)s + \cdots \tag{5.130}$$

The dominant process contributing to $W_{\mu\nu}$ is the handbag diagram of Fig. 5.10 and gives

$$F_2^{eN} = x \sum_q e_q^2 \big[q(x) + \bar{q}(x)\big] \tag{5.131}$$

where the quark distribution functions are the expectations of quarks having fractional momentum $x$ to $x \to \delta x$; that is,

$$\langle N|\bar{q}q|N\rangle = q_N(x) + \bar{q}_N(x) \tag{5.132}$$

Thus, for the proton one has

$$F_2^{ep}(x) = \frac{4}{9}x\big[u_p(x) + \bar{u}_p(x)\big] + \frac{1}{9}x\big(d_p(x) + \bar{d}_p(x)\big) + \cdots \tag{5.133}$$

while for the neutron

$$F_2^{ep}(x) = \frac{4}{9}x\left[u_n(x) + \bar{u}_n(x)\right] + \frac{1}{9}x\left(d_n(x) + \bar{d}_n(x)\right) + \cdots \tag{5.134}$$

From isospin invariance, one expects that

$$u_p(x) = d_n(x) \tag{5.135}$$

$$\bar{u}_p(x) = \bar{d}_n(x) \tag{5.136}$$

$$d_p(x) = u_n(x) \tag{5.137}$$

$$\bar{d}_p(x) = \bar{u}_n(x) \tag{5.138}$$

For spin-$\frac{1}{2}$ quarks one can derive the Callan–Gross relation [63] that

$$F_2(x) = 2xF_1(x) \tag{5.139}$$

to be compared to $F_1 = 0$ for scalar quarks.

For neutrino scattering, one has

$$F_2^{\nu p}(x) = F_2^{\bar{\nu} n}(x) = 2x\left[d_p(x) + \bar{u}_p(x)\right] + \cdots \tag{5.140}$$

$$F_2^{\nu n}(x) = F_2^{\bar{\nu} p}(x) = 2x\left[u_p(x) + \bar{d}_p(x)\right] + \cdots \tag{5.141}$$

while

$$F_3^{\nu p}(x) = F_3^{\bar{\nu} n}(x) = 2x\left[\bar{u}_p(x) - d_p(x)\right] + \cdots \tag{5.142}$$

$$F_3^{\bar{\nu} p}(x) = F_3^{\nu n}(x) = 2x\left[\bar{d}_p(x) - u_p(x)\right] + \cdots \tag{5.143}$$

Clearly, the quark distributions must be normalized such that

$$\int_0^1 dx \begin{pmatrix} u_p(x) - \bar{u}_p(x) \\ d_p(x) - \bar{d}_p(x) \\ s_p(x) - \bar{s}_p(x) \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \tag{5.144}$$

A very important result of the experimental analysis concerns the total fraction of momentum carried by quarks and antiquarks in the nucleus:

$$\sum_q \int_0^1 dx\, x\left[q(x) + \bar{q}(x)\right] = 1 - \epsilon \tag{5.145}$$

the result is [64–66] that $\epsilon \approx 0.5$; we interpret this as about half the proton momentum being carried by gluons. There are other important sum rules, such as the Adler sum rule [67],

$$\int_0^1 \frac{dx}{2x}\left[F_2^{\bar{\nu}p}(x) + F_2^{\nu p}(x)\right] = 1 \tag{5.146}$$

following from current algebra assumptions; similar considerations lead to the Gross–Llewellyn Smith sum rule [68], which states that

$$\int_0^1 dx\left[F_3^{\nu p}(x) + F_3^{\bar{\nu}p}(x)\right] = -6 \tag{5.147}$$

The usual fractional-charge quarks allow

$$\frac{1}{4} \leqslant \frac{F_2^{en}}{F_2^{ep}} \leqslant 4 \tag{5.148}$$

while Han–Nambu integer charges [69] would have required

$$\frac{F_2^{en}}{F_2^{ep}} \geqslant \frac{1}{2} \tag{5.149}$$

which seems to be ruled out near $x = 1$.

Comparison of weak and electromagnetic exchanges dictates that

$$F_3^{\nu p} - F_3^{\bar{\nu}p} = 12\left(F_1^{ep} - F_1^{en}\right) \tag{5.150}$$

$$F_2^{\nu p} + F_2^{\nu n} = \frac{18}{5}\left(F_2^{ep} + F_2^{en}\right) \tag{5.151}$$

and these strikingly confirm the expectation that the same quark substructure is probed by the photon and $W^\pm$ bosons.

The quark distribution functions themselves should be calculable from QCD, perhaps by lattice methods. It is amusing that certain features of the $x$-dependence can be understood in terms of Regge poles. Experimental measurements indicate that $F_1(x) \sim 1/x$ and $F_2(x) \sim$ constant as $x \to 0$. A Regge pole will give Fig. 5.11

$$W_1\left(\nu, Q^2\right) \sim \nu^\alpha f_1\left(Q^2\right) \tag{5.152}$$

$$\nu W_2\left(\nu, Q^2\right) \sim \nu^{\alpha-1} f_2\left(Q^2\right) \tag{5.153}$$

so Bjorken scaling requires that $f_1 \sim (Q^2)^{-\alpha}$ and $f_1 \sim (Q^2)^{1-\alpha}$, whereupon $F_1 \sim x^{-\alpha}$ and $F_2 \sim x^{1-\alpha}$. Thus phenomenology suggests that $\alpha = 1$ near $x \sim 0$; this is reasonable and says that the quark sea of $q\bar{q}$ pairs dominates at small $x$ since it is flavor singlet and experiences Pomeron ($\alpha = 1$) exchange. (Such a sepa-

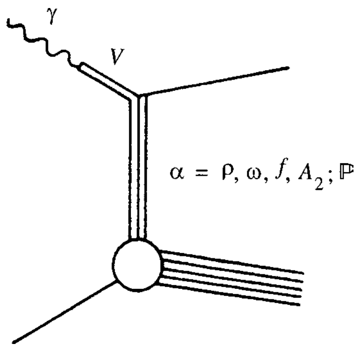



**Figure 5.11** Regge pole interpretation of $x$ dependencies.

ration of diffractive sea and normal Regge valence quarks was used in, for example, Ref. [70].)

This completes a brief overview of the naive model with exact scaling. Perturbative QCD makes no claim to calculate the $x$-dependence; the point at issue is that Bjorken scaling is strict only for $Q^2 \to \infty$ and, in fact, the scaling functions do exhibit some departure and do depend on $Q^2$: $F_i(x, Q^2)$ $(i = 1, 2, 3)$. Asymptotic freedom not only underwrites the scaling limit but allows calculation of the *scaling violations* [i.e., the logarithmic $Q^2$ dependence of $F_i(x, Q^2)$].

The point is that QCD insists that in addition to the diagram of Fig. 5.12*a*, there are the inevitable corrections of Fig. 5.12*b* and higher orders. Even though, as it will turn out, we can presently check these QCD corrections quantitatively only at a 10% level, we should recall that the nature of the $Q^2 \to \infty$ limit already uniquely distinguishes QCD as the correct theory [19].

To apply the renormalization group to the deep Euclidean region, it is convenient to use the operator product expansion [11] on the light cone [71–74]. The Fourier transform

$$W_{\mu\nu} = \frac{1}{4\pi} \int d^4z\, e^{iq\cdot z} \langle p | J_\mu^+(z) J_\nu(0) | p \rangle \tag{5.154}$$

is dominated, when $|q^2| \to \infty$ by the neighborhood of $z^2 \sim 0$. We wish to predict, in particular, the momentum dependence of *moments* of structure functions in the form

$$\int_0^1 dx\, x^{n-2} F_2(x, Q^2) = C_n(Q^2) \langle N | 0^n | N \rangle \tag{5.155}$$

where the large $Q^2$ dependence of $C_n$ will be determined by QCD and the renormalization, while the matrix element is related to the quark distribution functions to be fitted to experiment. The $C_n$ and $O^n$ are often called [11] Wilson coefficients and Wilson operators, respectively. For $z^2 \approx 0$ we thus write

$$A(z)B(0) \simeq \sum_{i,n} C_i^n(z^2) z_{\mu_1} z_{\mu_2} \cdots z_{\mu_n} O_i^{\mu_1\mu_2\cdots\mu_n}(0) \tag{5.156}$$

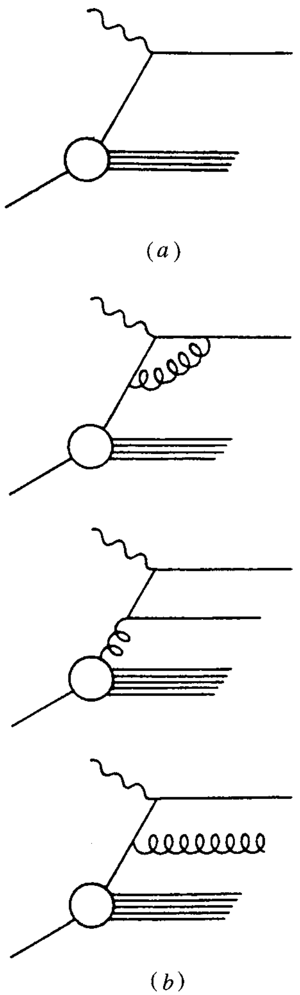



**Figure 5.12** QCD corrections to parton model.

where the sum is over spin ($n$) and operator types ($i$). The operators are symmetric and traceless in the indices $\mu$. As $z^2 \to 0$, the Wilson coefficients behave as

$$C_i^n(z^2) \underset{z^2 \to 0}{\sim} \left(\frac{1}{z^2}\right)^{(d_A + d_B + n - d_{0_i})/2} \tag{5.157}$$

where $d_A$, $d_B$, and $d_0$ are the naive dimensions. For given $A$ and $B$ the strongest singularity on the light cone will be from the term in Eq. (5.156) of minimum twist ($\tau$) defined by the difference of dimension and spin:

$$\rho = d_{0_i} - n \tag{5.158}$$

To give examples, for a scalar field $\phi$ the following operators have twist $\tau = 1$:

$$\phi,\ \partial_\mu \phi,\ \partial_\mu \partial_\nu \phi \tag{5.159}$$

Of course, these are irrelevant to deep-inelastic scattering where the fields are quarks $\psi$ and gluons $A_\mu$ and the lowest twist is $\tau = 2$ for the two types of operator,

$$O_F^n = \frac{1}{n!}\left[\bar{\psi}\gamma_{\mu_1} D_{\mu_2} \cdots D_{\mu_n} \begin{pmatrix} \lambda_a \\ 1 \end{pmatrix} \psi + \cdots \text{permutations}\right] \tag{5.160}$$

$$O_V^n = \frac{1}{n!}(F_{\alpha\mu_1} D_{\mu_2} D_{\mu_3} \cdots D_{\mu_{n-1}} F_{\alpha\mu_n} + \cdots \text{permutations}) \tag{5.161}$$

where in $O_F^n$ the $\lambda_a$ matrix is in flavor space, and its replacement by 1 is the flavor-singlet case. For flavor nonsinglet there is only one type of operator ($O_F$), while for flavor singlet there are two ($O_F$ and $O_V$). As we shall see, the consequent operator mixing makes moments of the flavor-singlet structure functions somewhat more complicated to analyze.

The first step is to go into momentum space by the Fourier transform (suppressing two obvious Lorentz indices):

$$\int d^4z\, e^{iq\cdot z} \langle p | J(z) J(0) | p \rangle$$

$$\underset{z^2 \to 0}{\approx} \sum_{i,n} \int d^4z\, e^{iq\cdot z} C_i^n(z^2) z_{\mu_1} z_{\mu_2} \cdots z_{\mu_n} \langle p | Q_i^{\mu_1\mu_2\cdots\mu_n} | p \rangle \tag{5.162}$$

$$= \sum_{i,n} (Q^2)^{-n} q_{\mu_1} q_{\mu_2} \cdots q_{\mu_n} C_i^n(Q^2) \langle p | Q_i^{\mu_1\mu_2\cdots\mu_n} | p \rangle \tag{5.163}$$

Next we expand the matrix element of the Wilson operator according to

$$\langle p | Q_i^{\mu_1\mu_2\cdots\mu_n} | p \rangle = A_i^n (p^{\mu_1} p^{\mu_2} \cdots p^{\mu_n} - m^2 g^{\mu_1\mu_2\mu_3\mu_4} \cdots p^{\mu_n} \cdots) \tag{5.164}$$

where trace terms are subtracted to make the tensor symmetric and traceless, and hence of definite spin. In terms of the scaling variable $x = (Q^2/2p \cdot q)$, this gives an expansion

$$\sum_{i,n} C_i^n(Q^2) x^{-n} A_i^n + O\left(x^{-n+2}\frac{m^2}{Q^2}\right) \tag{5.165}$$

To calculate the scaling violations at leading order, one will omit *both* (1) the trace terms, which give quark mass effects, and (2) higher-twist $\tau \geqslant 4$ operators, which give $m^2/Q^2$ corrections that are more challenging to calculate.

Now any physically measurable quantity must satisfy a renormalization group equation to exhibit independence of the scheme chosen. In particular, the matrix elements must satisfy

$$\left(\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}+2\gamma_\phi\right)\langle\phi|(JJ)|\phi\rangle=0 \tag{5.166}$$

where $|\phi\rangle$ may be any basic state of QCD (i.e., either a quark or a gluon state). In Eq. (5.166) we substituted the anomalous dimension $\gamma_J = 0$. It follows that for each $O_i^n$ we must have

$$\left(\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}+2\gamma_\phi\right)C_i^n\left(\frac{Q^2}{\mu^2, g(\mu)}\right)\langle\phi|O_i^n|\phi\rangle=0 \tag{5.167}$$

since each $O_i^n$ has a different and independent tensor structure. But the matrix element in Eq. (5.167) is itself a physically measurable quantity corresponding to a parton distribution function and hence

$$\left(\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}+2\gamma_\phi+\gamma_{O_i^n}\right)\langle\phi|O_i^n|\phi\rangle=0 \tag{5.168}$$

and it follows that

$$\left(\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}-\gamma_{O_i^n}\right)C_i^n\left(\frac{Q^2}{\mu^2, g(\mu)}\right)=0 \tag{5.169}$$

The method to calculate $\gamma_{O_i^n}$ is already evident from Eq. (5.168) since if we normalize at $p^2 = -\mu^2$ such that

$$\langle\phi|O_i^n|\phi\rangle=1 \quad \left(p^2=-\mu^2\right) \tag{5.170}$$

and then radiative corrections give for the leading logarithmic correction

$$\langle\phi|O_i^n|\phi\rangle=1+g^2b_i^n\ln\left(\frac{|p^2|}{\mu^2}\right)+O\left(g^4\right) \tag{5.171}$$

we have from Eq. (5.168)

$$\gamma_{O_i^n}=2g^2b_i^n-2\gamma_\phi \tag{5.172}$$

The solution of the renormalization group equation for $C_i^n$, Eq. (5.169), is given by

$$C_i^n\left(\frac{Q^2}{\mu^2, g(\mu^2)}\right)=C_i^n\left(1,\bar{g}\left(\mu^2\right)\right)\exp\left\{\int_0^{1/2\ln Q^2/\mu^2}dt'\gamma_{O_i^n}\left[\bar{g}(t')\right]\right\} \tag{5.173}$$

where the exponential sums the leading-logarithmic corrections to all orders of perturbation theory, as usual.

If there is more than one operator $O_i^n$ with the same Lorentz structure and internal quantum numbers, the $O_i^n$ are not separately multiplicatively renormalizable. Then we must replace Eq. (5.169) by the matrix equation

$$\sum_j \left[ \delta_{ij} \left( \mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial g} \right) - \gamma_{ij}^n \right] C_j^n \left( \frac{Q^2}{\mu^2, g(\mu)} \right) = 0 \tag{5.174}$$

and the matrix $\gamma_{ij}^n$ must be diagonalized before solving, as in Eq. (5.173) for the mixed operators. For the leading twist $\tau = 2$ flavor-nonsinglet case, there is only one operator type, given in Eq. (5.160), so we can solve immediately as in Eq. (5.173).

Going back to Eq. (5.165), we explain now why it is the moments of structure functions for which the $Q^2$ dependence is most directly provided by the renormalization group. From Eqs. (5.162) through (5.165) we have that the transition $T$-matrix for virtual Compton scattering has the leading-order expansion

$$T(x, Q^2) = \sum_{i,n} C_i^n \left( \frac{Q^2}{\mu^2, g(\mu)} \right) x^{-n} A_i^n \tag{5.175}$$

In the complex $x$-plane, $T$ is analytic apart from the physical cut along the real axis for $-1 \leqslant x \leqslant +1$. Encircling this cut by a Cauchy contour $C$ gives

$$\frac{1}{2\pi i} \int_C dx\, x^{n-2} T(x, Q^2) = \sum_i C_i^n \left( \frac{Q^2}{\mu^2, g(\mu)} \right) A_i^n \tag{5.176}$$

Shrinking $C$ to the physical cut gives the discontinuity formula required and by an obvious generalization of the optical theorem one obtains for a generic structure function $F(x, Q^2)$

$$\int_0^1 dx\, x^{n-2} F(x, Q^2) = \sum_i C_i^n \left( \frac{Q^2}{\mu^2, g(\mu)} \right) A_i^n \tag{5.177}$$

as promised earlier. Here $F$ may be $xF_1$, $F_2$, or $xF_3$. Adopting the convenient shorthand that

$$\langle F(Q^2) \rangle_n = \int_0^1 dx\, x^{n-2} F(x, Q^2) \tag{5.178}$$

and using Eq. (5.173), we see that, normalizing to some $Q^2 = Q_0^2$, we have

$$\langle F(Q^2) \rangle_n = \langle F(Q_0^2) \rangle_n \exp\left[ - \int_{1/2 \ln Q_0^2/\mu^2}^{1/2 \ln Q^2/\mu^2} dt' \gamma_{O_i^n} \left( \bar{g}(t') \right) \right] \tag{5.179}$$

In lowest order, $\beta = -bg^3$ and $\gamma_{O_i^n} = C_i^n g^2$, and hence

$$\frac{\langle F_i(Q^2)\rangle_n}{\langle F_i(Q_0^2)\rangle_n} = \left[\frac{\alpha_s(Q_0^2)}{\alpha_s(Q^2)}\right]^{-C_i^n/2b} \tag{5.180}$$

$$= \exp(-a_i^n s) \tag{5.181}$$

where we used

$$\alpha_s = \frac{1}{4\pi b \ln(Q^2/\Lambda^2)} \tag{5.182}$$

and defined

$$s = \ln\frac{\ln(Q^2/\Lambda^2)}{\ln(Q_0^2/\Lambda^2)} \tag{5.183}$$

$$a_i^n = \frac{C^n}{2b} = \frac{\gamma_{O_i^n}}{2bg^2} = \frac{\gamma_{O_i^n}}{8\pi b\alpha_s} \tag{5.184}$$

We can delay no longer calculation of at least the flavor-nonsinglet anomalous dimension for the operator $O_F^n$ of twist 2 given in Eq. (5.160). In particular, we wish to calculate $\gamma_{FF}^F$ for the matrix element of $O_F^n$ between fermion (quark) states and then compute $\gamma_{O_F^n}$ by using Eq. (5.172). To accomplish this, we need the Feynman rules for the insertion of $O_F^n$; these are indicated in Fig. 5.13, where $\Delta_\mu$ is an arbitrary fixed four-vector. Let $Z = Z(\Lambda/\mu, g)$ be the multiplicative renormalization constant for $O_F^n$; then the anomalous dimension is

$$\gamma_{O_F^n} = \mu\frac{\partial}{\partial\mu}\ln Z\left(\frac{\Lambda}{\mu, g}\right) = -\frac{\partial}{\partial(\ln\Lambda)}\ln Z\left(\frac{\Lambda}{\mu, g}\right) \tag{5.185}$$

The relevant Feynman diagrams are shown in Fig. 5.14. Let us for clarity write

$$Z = 1 + (c_1 + c_2 + c_3)\ln\Lambda + \text{two-loops} \tag{5.186}$$

where $c_1$, $c_2$, and $c_3$ refer, respectively, to the contributions of Fig. 5.14*a* through *c*.

To proceed with the calculation, we first state and prove by induction the following identity:

$$\frac{1}{2\pi^2}\int d\Omega\, \hat{k}_{\mu_1}\hat{k}_{\mu_2}\cdots\hat{k}_{\mu_n}$$

$$= \begin{cases} \frac{2}{(n+2)!!}\sum_{\text{pairs}} g_{\mu_1\mu_2}g_{\mu_3\mu_4}\cdots g_{\mu_{n-1}\mu_n} & (n \text{ even}) \\ 0 & (n \text{ odd}) \end{cases} \tag{5.187}$$

Here the integral is over the three-dimensional surface of a unit four-hyper-sphere.

For $n$ odd, $k_\mu \to -k_\mu$ merely interchanges hyperhemispheres but changes the overall sign. Hence the integral vanishes.

For $n$ even, let us denote the correct normalization of the right-hand side by $N(n)$; it is then required to prove by induction that $N(n) = 2/(n+2)!!$

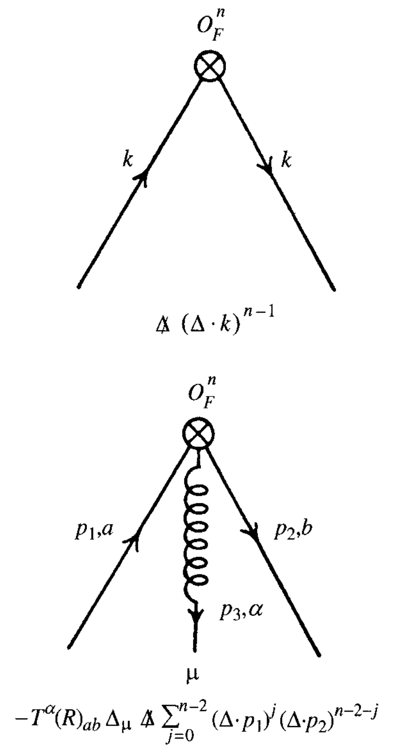



**Figure 5.13** Feynman rules for insertion of operator $O_F^n$.

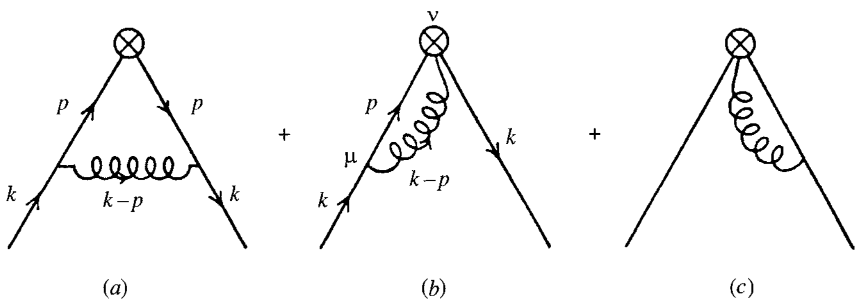



**Figure 5.14** Feynman diagrams used to compute anomalous dimension of $O_F^n$.

For $n = 0$, $N(0) = 1$ (surface area). Also by inspection $N(2) = \frac{1}{4}$ (contraction with $g^{\mu^1\mu^2}$) and $N(4) = \frac{1}{24}$ (contraction with $g^{\mu^1\mu^2}g^{\mu^3\mu^4}$).

Suppose that the required result is correct for $N(n)$, then consider $N(n+2)$. The sum over pairs now has $\nu(n+2)$ terms, where

$$\nu(n+2) = (n+1)\nu(n) \tag{5.188}$$

as can be seen from counting pairs

$$\nu(n) = \binom{n}{2}\binom{n-2}{2}\cdots\binom{2}{2}\frac{1}{(n/2)!} = \frac{n!}{n!!} \tag{5.189}$$

Now consider the expression for $(n+2)$ and contract with $g^{\mu_{n+1}\mu_{n+2}}$. There are $\nu(n)$ terms containing $g_{\mu_{n+1}\mu_{n+2}}$ and hence giving a factor of 4; the other $n\nu(n)$ terms do not contain $g_{\mu_{n+1}\mu_{n+2}}$ and give a factor of 1. Thus

$$N(n) = (4+n)N(n+2) \tag{5.190}$$

But we are assuming that $N(n) = 2/(n+2)!!$ and Eq. (5.190) then gives $N(n+2) = 2/(n+4)!!$. Since the formula is correct for $n = 0$, it is true for all even $n$. Q.E.D.

The Feynman rules for Fig. 5.14$a$ give the amplitude

$$ig^2C_2(R)\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^2(p-k)^2}(\Delta\cdot p)^{n-1}\gamma_\mu \not{p}\not{\Delta}\not{p}\gamma_\mu \tag{5.191}$$

where $\mathrm{Tr}[T^\alpha(R)T^\beta(R)] = \delta^{\alpha\beta}C_2(R)$ and we shall work in Feynman gauge ($\alpha = 1$) since at lowest order the anomalous dimension is gauge independent.

Hence

$$c_1(\Delta\cdot k)^{n-1}\not{\Delta}\ln\Lambda \approx 2ig^2C_2(R)\int\frac{d^4p}{(2\pi)^4}(\Delta\cdot p)^{n-1}\frac{p^2\not{\Delta}-2(p\cdot\Delta)\not{p}}{p^2(p-k)^2} \tag{5.192}$$

in which it is understood to mean the log divergence of the right-hand integral is to be taken.

Now differentiate Eq. (5.192) $(n-1)$ times with respect to $k_\mu$ by acting with

$$D_{n-1}^{\epsilon_1\epsilon_2\cdots\epsilon_{n-1}} = \frac{d^{n-1}}{dk_{\epsilon_1}dk_{\epsilon_2}\cdots dk_{\epsilon_{n-1}}} \tag{5.193}$$

On the left-hand side,

$$D_{n-1}(\Delta\cdot k)^{n-1} = (n-1)!\Delta_{\epsilon_1}\Delta_{\epsilon_2}\cdots\Delta_{\epsilon_{n-1}} \tag{5.194}$$

while on the right-hand side the log divergence is from

$$D_{n-1}\frac{1}{(p-k)^2} \approx (2n-2)!!\frac{p_{\epsilon_1}p_{\epsilon_2}\cdots p_{\epsilon_{n-1}}}{(p^2)^n} \tag{5.195}$$

The right-hand side of Eq. (5.192) is, therefore, after rotating to Euclidean space,

$$g^2C_2(R)\Delta_{\delta_1}\Delta_{\delta_2}\cdots\Delta_{\delta_{n-1}}\ln\Lambda(2-2)!!$$
$$\cdot\int\frac{d\Omega}{(2\pi)^4}\hat{p}_{\delta_1}\hat{p}_{\delta_2}\cdots\hat{p}_{\delta_{n-1}}(\not{\Delta}-2\hat{p}_{\delta_n}\gamma_{\epsilon_n}\hat{p}_{\epsilon_n}\Delta_{\delta_n})\hat{p}_{\epsilon_1}\hat{p}_{\epsilon_2}\cdots\hat{p}_{\epsilon_{n-1}} \tag{5.196}$$

Now using the identity, Eq. (5.187), enables us to evaluate this integral to find (choosing $\Delta^2 = 0$ is most convenient)

$$g^2 C_2(R) \Delta_{\delta_1} \Delta_{\delta_2} \cdots \Delta_{\delta_{n-1}} (\ln \Lambda)(2n-2)!! \frac{2\pi^2 \not{\Delta}}{(2\pi)^4} \cdot \left[ (n-1)! \frac{2}{2n!!} - (n-1)(n-1)! \frac{4}{(2n+2)!!} \right] \tag{5.197}$$

and hence

$$c_1 = +\frac{\alpha_s}{2\pi} C_2(R) \frac{2}{n(n+1)} \tag{5.198}$$

Next we compute $c_2$ from Fig. 5.14*b*. Again in Feynman gauge ($\alpha = 1$) we find that

$$c_2 (\Delta \cdot k)^{n-1} \not{\Delta} \ln \Lambda \approx -i g^2 C_2(R) \cdot \int \frac{d^4 p}{(2\pi)^4} \frac{\gamma_\mu \not{p} \Delta_\mu \not{\Delta}}{p^2 (p-k)^2} \sum_{j=0}^{n-2} (\Delta \cdot p)^j (\Delta \cdot k)^{n-2-j} \tag{5.199}$$

Using $\Delta^2 = 0$ we may rewrite $\not{\Delta}\not{p}\not{\Delta} = 2\not{\Delta}\Delta \cdot p$. Now act with $D_{n-1}$ of Eq. (5.193) on both sides of Eq. (5.199). As before, the left-hand side uses Eq. (5.194). The right-hand side is

$$-g^2 C_2(R) 2\not{\Delta} \int \frac{d^4 p}{(2\pi)^4} \frac{p \cdot \Delta}{p^2} \sum_{j=0}^{n-2} (\Delta \cdot p)^j D_{n-1} \frac{(\Delta \cdot k)^{n-2-j}}{(p-k)^2} \tag{5.200}$$

Now use, for the log divergent piece,

$$D_{n-1}^{\epsilon_1 \cdots \epsilon_{n-1}} \frac{(\Delta \cdot k)^{n-2-j}}{(p-k)^2} \approx \binom{n-1}{j+1} \frac{(n-j-2)!(2j+2)!!}{(p^2)^j} \cdot p_{\epsilon_1} p_{\epsilon_2} \cdots p_{\epsilon_{j+1}} \Delta_{\epsilon_{j+2}} \Delta_{\epsilon_{j+3}} \cdots \Delta_{\epsilon_{n-1}} \tag{5.201}$$

to obtain

$$\begin{aligned} &-g^2 C_2(R) 2\not{\Delta} \int \frac{d^4 p}{(2\pi)^4} \sum_{j=0}^{n-2} \frac{(\Delta \cdot p)^{j+1}}{(p^2)^{j+1}} \frac{(n-1)!(2j+2)!!}{(j+1)!} \cdot p_{\epsilon_1} p_{\epsilon_2} \cdots p_{\epsilon_{j+1}} \Delta_{\epsilon_{j+2}} \cdots \Delta_{\epsilon_{n-1}} \\ &= -g^2 C_2(R) \not{\Delta} \ln \Lambda \sum_{j=0}^{n-2} \Delta_{\delta_1} \cdots \Delta_{\delta_{j+1}} \Delta_{\epsilon_{j+2}} \cdots \Delta_{\epsilon_{n-1}} (n-1)!(2j+2)!! \cdot \frac{1}{(2\pi)^4} \int d\Omega \hat{p}_{\delta_1} \cdots \hat{p}_{\delta_{j+1}} \hat{p}_{\epsilon_1} \cdots \hat{p}_{\epsilon_{j+1}} \end{aligned} \tag{5.202}$$

and, using identity (5.187), this becomes

$$\frac{-g^2}{8\pi^2} C_2(R) \not{\Delta} \Delta_{\epsilon_1} \cdots \Delta_{\epsilon_{n-1}} (n-1)! \sum_{j=0}^{n-2} (2j+2)!! \frac{2 \ln \Lambda}{(2j+4)!!} \tag{5.203}$$

and hence

$$c_2 = -\frac{\alpha_s C_2(R)}{2\pi} \sum_{j=2}^{n} \frac{2}{j} \tag{5.204}$$

It is easy to show that the graph of Fig. 5.14*c* gives the same result $c_3 = c_2$, and hence from Eqs. (5.186), (5.198), and (5.204),

$$\gamma_{O_F^n} = \frac{\alpha_s C_2(R)}{2\pi} \left[ -\frac{2}{n(n+1)} + 4 \sum_{j=2}^{n} \frac{1}{j} \right] \tag{5.205}$$

For the dimension $\gamma_{FF}^{F}$,

$$\gamma_{FF}^{F} = \gamma_{O_F^n} + 2\gamma_F \tag{5.206}$$

But we have

$$\gamma_F = -\frac{\partial}{\partial \ln \Lambda} \ln Z_F \left( \frac{\Lambda}{\mu, g} \right) \tag{5.207}$$

and from Fig. 5.15 (evaluated for the Abelian case of quantum electrodynamics in Chapter 3),

$$Z_F = 1 - \frac{\alpha_s}{4\pi} C_2(R) \ln \Lambda \tag{5.208}$$

whence

$$\gamma_{O_F^n} = \frac{\alpha_s C_2(R)}{2\pi} \left[ 1 - \frac{2}{n(n+1)} + 4 \sum_{j=2}^{n} \frac{1}{j} \right] \tag{5.209}$$

is the anomalous dimension to be used for the nonsinglet moments in Eq. (5.180).

We defer a detailed phenomenology but it is irresistible to show two of the early figures that exemplify the excellent *qualitative* support for perturbative QCD from experimental data. Present 2008 data more strongly confirm this support.

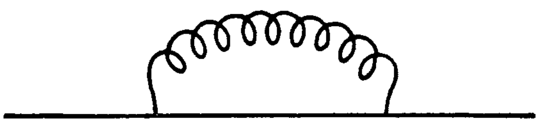


**Figure 5.15** Graph giving anomalous dimension $\gamma_F$.

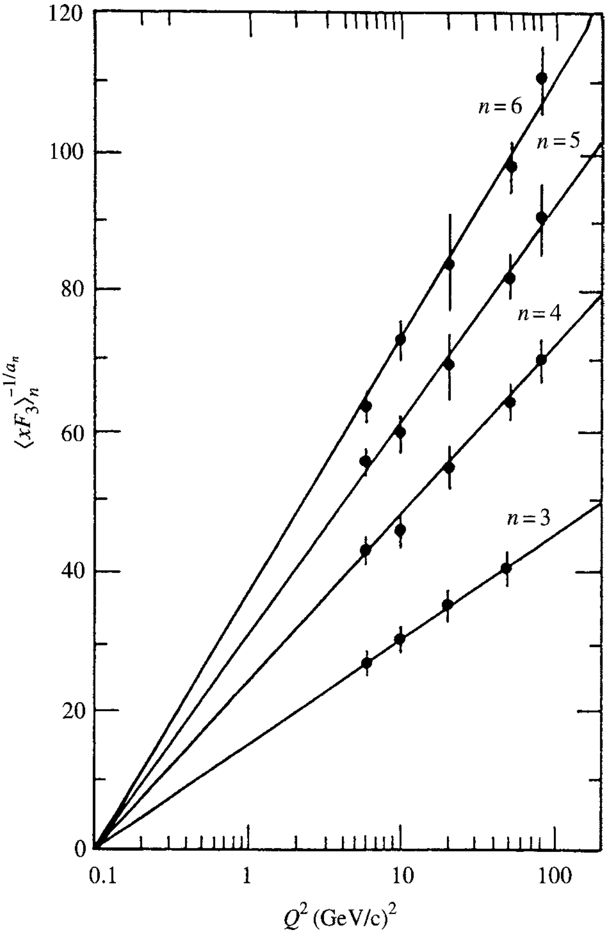



**Figure 5.16** Logarithmic $Q^2$ dependence of structure function moments. (Data from Ref. [75].)

In Eq. (5.180) the nonsinglet structure function may be taken as $xF_3^{\nu N}$. Then according to Eq. (5.180), a plot of the $n$th moment $\langle xF_3 \rangle_n$ raised to the power $(-1/a_n)$ should be linear in $\ln Q^2$. Experimental support for such logarithmic dependence of the scaling violations is shown in Fig. 5.16.

Further, we may compare the QCD prediction for the relative slopes of the different moments since if we write Eq. (5.209) as

$$\gamma_{O_F^n} = \frac{\alpha C_2(R)}{2\pi} \delta(n) \tag{5.210}$$

we find that $\delta(3) = 4.166$, $\delta(4) = 5.233$, $\delta(5) = 6.066$, $\delta(6) = 6.752$, and $\delta(7) = 7.335$. Thus from the basic formula (5.181) we see that

$$\frac{d \ln \langle F_{NS}(Q^2) \rangle_n}{d \ln \langle F_{NS}(Q^2) \rangle_{n'}} = \frac{\delta(n)}{\delta(n')} \tag{5.211}$$

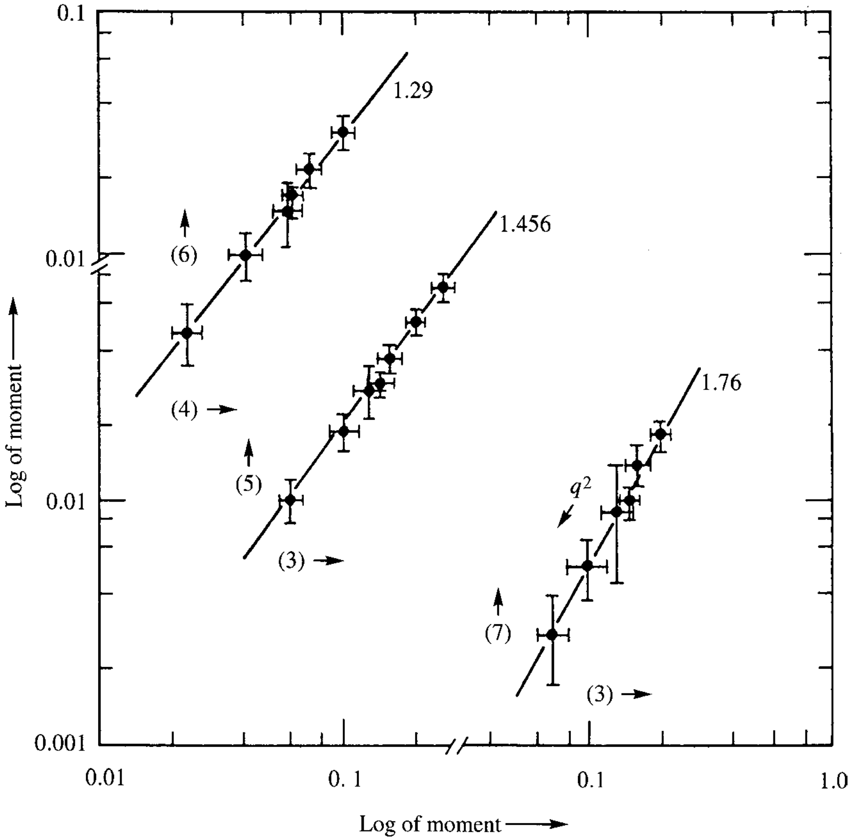



**Figure 5.17** Perkins plot checks $n$ dependence of anomalous dimension of $O_F^n$. (From Ref. [64].)

The resulting log-log plots are shown in Fig. 5.17. There is excellent qualitative agreement with the three ratios $\delta(6)/\delta(4) = 1.290$, $\delta(5)/\delta(3) = 1.456$, and $\delta(7)/\delta(3) = 1.760$. Figure 5.17 depicts a Perkins plot after an Oxford member of the collaboration.

The scaling violations offer good qualitative confirmation of QCD. Recall that quantum electrodynamics is confirmed to an accuracy of 1 in $10^6$ at order $\alpha^3$ in the electron anomalous magnetic moment (1 in $10^9$ in the magnetic moment itself); here QCD is tested at a 10% level. Nevertheless, the results are so encouraging and positive that in grand unification we shall assume QCD. Also, as we shall see, although scaling violations are a most important test, there are other independent checks of the theory.

To study flavor-singlet moments we need (in an obvious notation) $\gamma_{FF}^V$, $\gamma_{VV}^F$, $\gamma_{VV}^V$ as well as $\gamma_{FF}^F$, and then the matrix

$$\begin{pmatrix} \gamma_{FF}^F & \gamma_{FF}^V \\ \gamma_{VV}^F & \gamma_{VV}^V \end{pmatrix} \tag{5.212}$$

must be diagonalized to obtain simple predictions which are again supported by experimental data (e.g., Ref. [64]).

The anomalous dimensions were first computed by Gross and Wilczek [16, 17] and independently by Georgi and Politzer [76]. The origin of the various terms in, for example, Eq. (5.209) was a little mysterious when we computed the Feynman graphs. A more intuitive approach, based on parton (quark and gluon) splitting functions, was presented, for example, in an article by Altarelli and Parisi [77].

## 5.6 Background Field Method

Returning to formal considerations, we now introduce a procedure for quantization which keeps explicit gauge invariance. Normally, the classical Lagrangian is constructed to be gauge invariant, but on quantization the explicit gauge invariance is lost in the Feynman rules because of the necessity to add gauge fixing and Faddeev–Popov ghost terms. The background field method maintains explicit gauge invariance at the quantum level and hence is simpler both conceptually and technically. Here we shall introduce it only because of its technical superiority and lasting importance.

Historically, the method was introduced by DeWitt [78, 79], then by Honerkamp [80, 81], and was made popular by 't Hooft [82]. It is now being widely exploited [83–92] including its use in grand unification [88, 90], gravity [78, 79, 82, 91], and supergravity (e.g., Ref. [92]).

Here we shall derive the Feynman rules, then merely exemplify the method by calculating the $\beta$-function of the renormalization group. We follow closely the clear presentation of Abbott [87].

Let us begin with a nongauge theory of a scalar field $\Phi$ for which we write the generating functional

$$Z[J] = \int D\Phi \exp\left[i\left(S[\Phi] + \int d^4x J\cdot\Phi\right)\right] \tag{5.213}$$

The generating functional $W[J]$ for connected diagrams is

$$W[J] = -i \ln Z[J] \tag{5.214}$$

The effective action is then

$$\Gamma(\bar{\Phi}) = W[J] - J\cdot\bar{\Phi} \tag{5.215}$$

$$\bar{\Phi} = \frac{\delta W}{\delta J} \tag{5.216}$$

$\Phi(\bar{\phi})$ generates the proper, one-particle-irreducible diagrams.

Now let us decompose

$$\Phi = B + \phi \tag{5.217}$$

and quantize $\phi$ while leaving the background field $B$ unquantized. We generalize the definitions above to

$$\mathscr{Z}[j, B] = \int D\phi \exp\left\{i\left(S[B+\phi] + \int d^4x J \cdot \phi\right)\right\} \tag{5.218}$$

$$\mathscr{W}[J, B] = i \ln \mathscr{Z}[J, B] \tag{5.219}$$

$$\tilde{\Gamma}[\tilde{\phi}, B] = \mathscr{W}[J, B] - J \cdot \tilde{\phi} \tag{5.220}$$

$$\tilde{\phi} = \frac{\delta \mathscr{W}}{\delta J} \tag{5.221}$$

Why do this? The magic begins to appear when we substitute Eq. (5.217) into Eq. (5.218) to find that

$$\mathscr{Z}[J, B] = Z[J] \exp(-iJ \cdot B) \tag{5.222}$$

and hence

$$\mathscr{W}[J, B] = W[J] - J \cdot B \tag{5.223}$$

From Eqs. (5.216), (5.221), and (5.223) we deduce that

$$\tilde{\phi} = \bar{\Phi} - B \tag{5.224}$$

and substitution into Eq. (5.220) gives

$$\Gamma[\tilde{\phi}, B] = \Gamma[\tilde{\phi} + B] \tag{5.225}$$

This is an important result, especially the special case worth writing as a new equation

$$\Gamma[\tilde{\phi} = 0, B] = \Gamma[B] \tag{5.226}$$

Thus the normal effective action $\Gamma(B)$ where $B$ would be quantized can be computed by summing all vacuum graphs (no external $\phi$ lines) in the presence of the background $B$. This is already quite remarkable.

The two approaches are now:

1. Treat $B$ exactly and find the exact $\phi$ propagator in the presence of $B$. This is possible only for very simple $B$.
2. Treat $B$ perturbatively. Here we use normal $\phi$ propagators and introduce $B$ only on the external lines.

The second approach is used here.

In a gauge theory, we can use this procedure provided that we make a sufficiently clever choice of gauge. We write

$$\mathscr{Z}[J, B] = \int DA \det \mathscr{M}^{ab} \exp\left\{ iS[B + A] - \int d^4x \left( \frac{1}{2\alpha} \mathscr{G} \cdot \mathscr{G} + J \cdot A \right) \right\} \tag{5.227}$$

The appropriate (background) gauge is

$$\mathscr{G}^a = \partial_\mu A_\mu^a + g C^{abc} B_\mu^b A_\mu^c \tag{5.228}$$

This is the covariant derivative, with respect to $B_\mu^a$, acting on $A_\mu^a$. Now consider the infinitesimal transformation ($u^a$ = gauge function)

$$\delta B_\mu^a = -C^{abc} u^b B_\mu^c + \frac{1}{g} \partial_\mu u^a \tag{5.229}$$

$$\delta J_\mu^a = -C^{abc} u^b J_\mu^c \tag{5.230}$$

$$\delta A_\mu^a = -C^{abc} u^b A_\mu^c \tag{5.231}$$

It is not difficult to see that this leaves $\mathscr{Z}[J, B]$ of Eq. (5.227) invariant. First, $J \cdot A$ is invariant since

$$J' \cdot A' = (\mathbf{J} - \mathbf{u} \hat{} \mathbf{J}) \cdot (\mathbf{A} - \mathbf{u} \hat{} \mathbf{A}) \tag{5.232}$$

$$= \mathbf{J} \cdot \mathbf{A}^2 + O(u) \tag{5.233}$$

Next, since

$$\delta\left(B_\mu^a + A_\mu^a\right) = -\mathbf{u} \hat{} (\mathbf{B} + \mathbf{A}) + \frac{1}{g} \partial_\mu \mathbf{u} \tag{5.234}$$

is a gauge transformation, $S[B + A]$ is invariant. Because of the choice, Eq. (5.228), the adjoint rotation, Eq. (5.231), leaves $\mathscr{G}^a$ invariant. For similar reasons, the Faddeev–Popov determinant $\det \mathscr{M}^{ab}$ is invariant.

Hence $\mathscr{Z}[J, B]$ is invariant. Making the steps

$$\mathscr{W}[J, B] = -\ln \mathscr{Z}[J, B] \tag{5.235}$$

$$\Gamma[\mathscr{A}, B] = \mathscr{W}[J, B] - J \cdot \mathscr{A} \tag{5.236}$$

$$\mathscr{A} = \frac{\delta \mathscr{W}}{\delta J} \tag{5.237}$$

one finds that, in background field gauge, the effective action

$$\Gamma[\mathscr{A} = 0, B] = \Gamma[B] \tag{5.238}$$

is gauge invariant.

To compute this gauge-invariant action, we need to establish the Feynman rules to compute all 1PI diagrams with $B$ appearing only on external lines. The Faddeev–Popov ghosts term can be computed from Eqs. (5.228), (5.229), and (5.231): as

$$\mathscr{M}^{ab} = g\frac{\delta \mathscr{G}^a}{\delta u^b} = \Box\delta^{ab} - g\bar{\partial}_\mu C^{acb}\left(B^c_\mu + A^c_\mu\right) \\ + gC^{acb}B^c_\mu\bar{\partial}_\mu + g^2C^{ace}C^{edb}B^c_\mu\left(B^d_\mu + A^d_\mu\right) \tag{5.239}$$

and thence

$$\mathscr{S}_{\text{FPG}} = c^{a\dagger}M^{ab}c^b \tag{5.240}$$

The internal propagators and totally internal vertices are given by the old Feynman rules of Fig. 2.6. These must now be augmented by the external vertices displayed in Fig. 5.18.

This completes quantization, so let us discuss renormalization, which is our main concern. It is actually redundant to renormalize

$$c^a = \sqrt{\mathscr{Z}}c^a_R \tag{5.241}$$

$$A^a_\mu = \sqrt{Z_A}A^a_{\mu R} \tag{5.242}$$

since each internal propagator carries a $\sqrt{Z}$ at each end due to field renormalization and a $1/Z$ from propagator renormalization. We do need

$$g = Z_g g_R \tag{5.243}$$

$$B = \sqrt{Z_B}B_R \tag{5.244}$$

But $Z_g$ and $Z_B$ have the simple relation

$$Z_g\sqrt{Z_B} = 1 \tag{5.245}$$

as can be seen immediately by renormalization of the gauge-invariant quantity

$$F^a_{\mu\nu} = \partial_\mu B^a_\nu - \partial_\nu B^a_\mu + gC^{abc}B^b_\mu B^c_\nu \tag{5.246}$$

From Eq. (5.243) we have that

$$\mu\frac{\partial}{\partial\mu}g = 0 = Z_g\mu\frac{\partial}{\partial\mu}g_R + g_R\mu\frac{\partial}{\partial\mu}Z_g \tag{5.247}$$

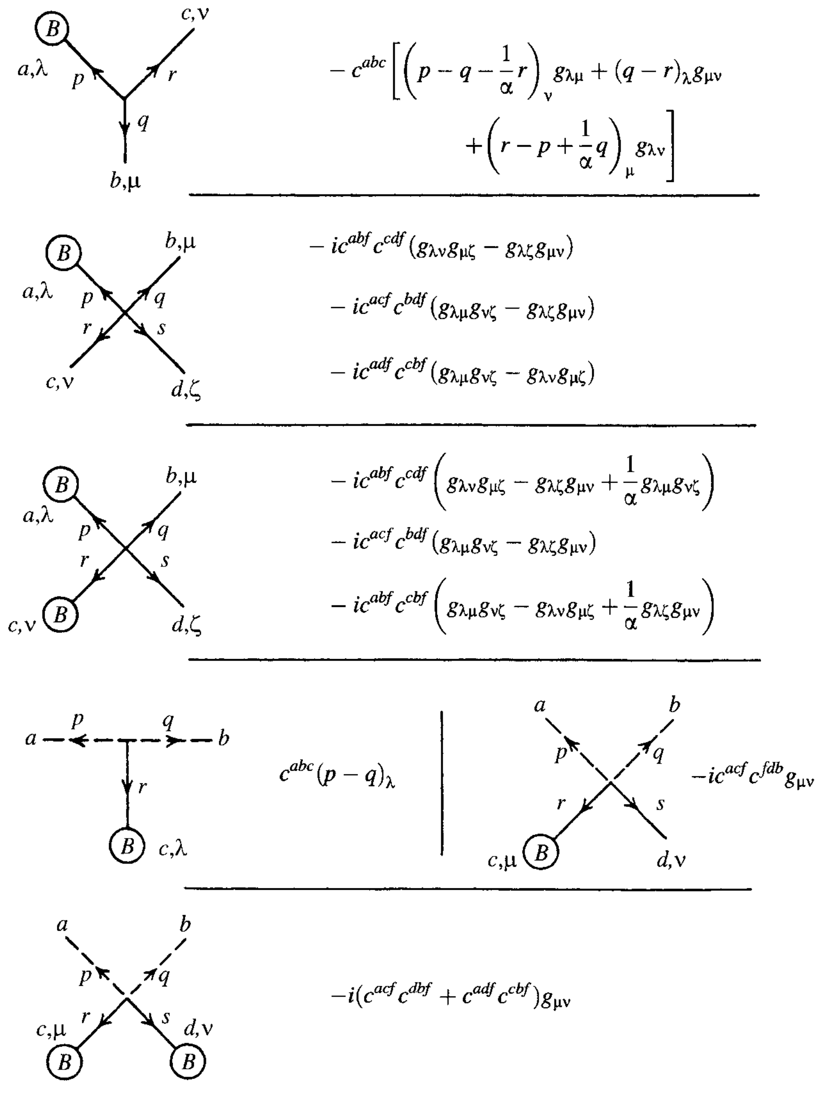



**Figure 5.18** Feynman rules in background field gauge. These rules involving external $B_\mu^a$ fields augment the rules given in Fig. 2.6.

and hence, as usual,

$$\beta = \mu\frac{\partial}{\partial\mu}g_R = -g_R\mu\frac{\partial}{\partial\mu}\ln Z_g \tag{5.248}$$

But because of Eq. (5.245), this becomes

$$\beta = \frac{1}{2}g_R\mu\frac{\partial}{\partial\mu}\ln Z_B = g_R\gamma_B \tag{5.249}$$

where $\gamma_B$ is the anomalous dimension of $B$.

Let us use dimensional regularization, defining dimension $d = (4 - 2\delta)$ and expand:

$$Z_B = 1 + \sum_{i=1}^{\infty} \frac{Z_B^{(i)}}{\delta^i} \tag{5.250}$$

Since, in a generic dimension, we must redefine $g \to \mu^{-\delta} g$, we see that (now we replace $g_R$ simply by $g$ from here on)

$$\mu \frac{\partial}{\partial \mu} g = -\delta g + \beta \tag{5.251}$$

Substitution of Eq. (5.250) into Eq. (5.249) gives

$$2 Z_B \gamma_B = \mu \frac{\partial}{\partial \mu} \left[ 1 + \sum_{i=1}^{\infty} \frac{Z_B^{(i)}}{\delta^i} \right] \tag{5.252}$$

Now use Eq. (5.251) to replace

$$\mu \frac{\partial}{\partial \mu} = \beta \frac{\partial}{\partial g} - \delta g \frac{\partial}{\partial g} \tag{5.253}$$

to discover that, by considering the coefficient of $\delta^{-i}$,

$$\beta \left( 2 - g \frac{\partial}{\partial g} \right) Z_B^{(i)} = -g^2 \frac{\partial}{\partial g} Z_B^{(i+1)} \tag{5.254}$$

In particular, since $Z_B^{(0)} = 1$, we have

$$\beta = -\frac{1}{2} g^2 \frac{\partial}{\partial f} Z_B^{(1)} \tag{5.255}$$

This sets up the calculation of $\beta$ in a simple form since once we evaluate the coefficient $\zeta$ in

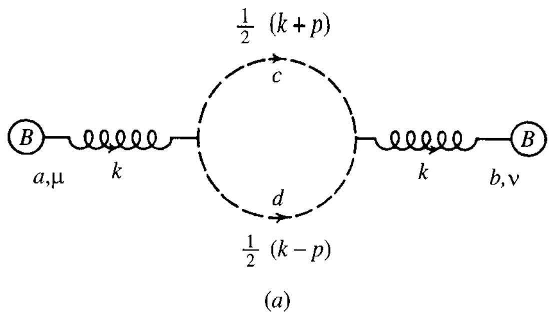
$$Z_B = 1 + \frac{\zeta}{\delta} \left( \frac{g}{4\pi} \right)^2 + O(g^4) \tag{5.256}$$

it follows that

$$\beta = \frac{-g\zeta}{2} \left( \frac{g}{4\pi} \right)^2 + O(g^5) \tag{5.257}$$

To find $\zeta$, only two diagrams are necessary, as shown in Fig. 5.19. Let us write

$$\zeta = \zeta_1 + \zeta_2 \tag{5.258}$$



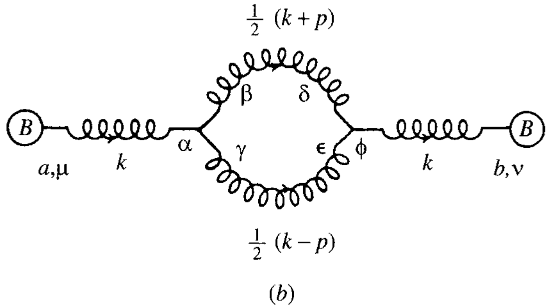




**Figure 5.19** Feynman diagrams needed to compute one-loop $\beta$-function in background gauge.

where $\zeta_1$ and $\zeta_2$ are the contributions, respectively, from Fig. 5.19*a* and *b*.

To compute $\zeta_1$ the amplitude corresponding to Fig. 5.19*a* is

$$-\int \frac{d^4p}{(2\pi)^4}\frac{1}{16}\cdot\frac{4i}{(k+p)^2}\frac{4i}{(k-p)^2}g^2C_{acd}C_{bdc}p_\alpha p_\beta$$

$$= \delta_{ab}C_2(G)\left(\frac{g}{4\pi}\right)^2\frac{1}{\pi^2}\int\frac{d^4p\, p_\alpha p_\beta}{(k+p)^2(k-p)^2} \tag{5.259}$$

The integral is

$$\int_0^1 dx\frac{d^4p\, p_\alpha p_\beta}{[p^2+2p\cdot k(2x-1)+k^2]^2}$$

$$= i\pi^2\ln\Lambda\int_0^1 dx\Big[k_\alpha k_\beta(2x-1)^2 2$$

$$+\frac{1}{2}g_{\alpha\beta}\left(-k^2(2x-1)^2+k^2\right)(-2)\Big] \tag{5.260}$$

$$= -i\pi^2\frac{2}{3}\left(k^2g_{\alpha\beta}-k_\alpha k_\beta\right) \tag{5.261}$$

and hence

$$\zeta_1 = \frac{2}{3} C_2(G) \tag{5.262}$$

Finally, Fig. 5.19*b* gives the amplitude (remembering a factor $\frac{1}{2}$ for the identical particles!). We put $\alpha = 1$ for simplicity; this diagram is separately gauge invariant.

$$\left(\frac{g}{4\pi}\right)^2 \frac{1}{\pi^2} C_2(G) I_{\mu\nu}(k) \tag{5.263}$$

$$I_{\mu\nu}(k) = \frac{1}{2}\int_0^1 \frac{dx\, d^4p}{[p^2 - 2p\cdot k(2x-1) + k^2]^2} \mathscr{I}_{\mu\nu}(k, p) \tag{5.264}$$

$$\mathscr{I}_{\mu\nu}(k, p) = \left[g_{\mu\beta}(-2k)_\gamma + g_{\beta\gamma}(p)_\mu + g_{\gamma\mu}(2k)_\beta\right]$$

$$\cdot \left[g_{\nu\delta}(2k)_\epsilon + g_{\delta\epsilon}(-p)_\nu + g_{\epsilon\nu}(-2k)_\delta\right] g_{\beta\delta} g_{\gamma\epsilon} \tag{5.265}$$

$$= 4\left(2k_\mu k_\nu - 2k^2 g_{\mu\nu} - p_\mu p_\nu\right) \tag{5.266}$$

Hence

$$I_{\mu\nu}(k) = -2\left(i\pi^2 \ln\Lambda\right)\left(\frac{10}{3}\right)\left(k^2 g_{\mu\nu} - k_\mu k_\nu\right) \tag{5.267}$$

and

$$\zeta_2 = \frac{20}{3} C_2(G) \tag{5.268}$$

The one-loop formula for the $\beta$-function is thus

$$\beta = -\frac{11 g^3}{48\pi^2} C_2(G) + O\left(g^5\right) \tag{5.269}$$

in agreement with Eq. (5.69). Note that only two-point functions were needed, in contrast to the earlier derivation in Section 5.3. Also, it is more agreeable to see the gauge invariance of each diagram here than the more inelegant restoration of gauge invariance by combination of different Feynman diagrams. The procedure has been continued to two loops (see Abbott [87]), and there the technical advantages become even more apparent.

## 5.7 Summary

We have seen already here very remarkable consequences of the renormalization group. It enables us to sum up, to all orders of perturbation theory, the leading logarithms from computation of only the logarithmic divergence of a one-loop graph. At first sight, this appears paradoxical since it seems that information is being created from nowhere. But it is crucial that a renormalizable theory is such that once we compute $Z$ at one loop, we control the leading divergences at all orders, so renormalizability itself correlates every order of perturbation theory. We may say that the "disease" of ultraviolet divergences is being converted into a virtue.

Asymptotic freedom identifies QCD as the unique theory of strong interactions, as well as placing a curious constraint on the number of quark flavors. Fortunately, this constraint appears to be comfortably satisfied at least in the energy region presently accessible experimentally it may well change at LHC energy. Further, we have seen how QCD predicts logarithmic scaling violations, and for the moments of the flavor-nonsinglet structure functions, these are qualitatively in agreement with experiment.

Grand unification is likewise underwritten by the renormalization group and it further supports the idea that a group such as the SU(5) or SO(10) could eventually unify strong and electroweak interactions, if the desert hypothesis is partially correct.

Finally, we have noted the technical advantage of using the background field method to compute the quantities appearing in the renormalization group equation.

## References


1. E.C.G. Stückelberg and A. Petermann, *Helv. Phys. Acta* 5, 499 (1953).
2. M. Gell-Mann and F.E. Low, *Phys. Rev.* 95, 1300 (1954).
3. L.V. Ovsiannikov, *Dokl. Akad. Nauk USSR* 109, 1112 (1956).
4. N.N. Bogoliubov and D.V. Shirkov, *Theory of Quantized Fields*, Interscience, New York, 1959.
5. L.P. Kadanoff, *Physics* 2, 263 (1966).
6. C.G. Callan, *Phys. Rev.* D2, 1541 (1970).
7. C.G. Callan, *Phys. Rev.* D5, 3202 (1972).
8. K. Symanzik, *Commun. Math. Phys.* 18, 227 (1970).
9. K. Symanzik, *Commun. Math. Phys.* 23, 49 (1971).
10. K. Symanzik, in *Springer Tracts in Modern Physics*, Vol. 57, G. Hohler, ed., Springer-Verlag, New York, 1971, p. 22.
11. K.G. Wilson, *Phys. Rev.* 179, 1499 (1969).
12. G. 't Hooft, unpublished remarks at Marseille Conference, 1972.
13. H.D. Politzer, *Phys. Rev. Lett.* 26, 1346 (1973).
14. H.D. Politzer, *Phys. Rep.* 14C, 129 (1974).
15. D.J. Gross and F. Wilczek, *Phys. Rev. Lett.* 26, 1343 (1973).
16. D.J. Gross and F. Wilczek, *Phys. Rev.* D8, 3633 (1973).
17. D.J. Gross and F. Wilczek, *Phys. Rev.* D9, 980 (1974).
18. A. Zee, *Phys. Rev.* D7, 3630 (1973).
19. S. Coleman and D.J. Gross, *Phys. Rev. Lett.* 31, 851 (1973).
20. H.M. Georgi, H.R. Quinn and S. Weinberg, *Phys. Rev. Lett.* 33, 451 (1974).
21. H. Georgi and S.L. Glashow, *Phys. Rev. Lett.* 32, 438 (1974).
22. S. Coleman, in *Proceedings of the International School of Physics, Ettore Majorana, Erice, 1971*, A. Zichichi, ed., Editrice Compositori, 1973, p. 359.
23. D.J. Gross, in *Methods in Field Theory (Les Houches 1975)*, R. Balian and J. Zinn-Justin, eds., North-Holland, Amsterdam, 1976, p. 144.
24. S.L. Adler, *Phys. Rev.* D5, 3021 (1972).
25. K.G. Wilson and J. Kogut, *Phys. Rep.* 12C, 75 (1974).
26. G. Altarelli, *Riv. Nuovo Cimento* 4, 335 (1974).

**27** A. Petennan, *Phys. Rep.* 53, 157 (1979).
**28** E. Reya, *Phys. Rep.* 69, 195 (1981).
**29** J.C. Taylor, *Nucl. Phys.* B33, 436 (1971).
**30** A.A. Slavnov, *Theor. Math. Phys.* 10, 99 (1972).
**31** J.C. Ward, *Phys. Rev.* 78, 182 (1950).
**32** Y. Takahashi, *Nuovo Cimento* 6, 371 (1957).
**33** L. Banyai, S. Marculescu and T. Vescan, *Nuovo Cimento Lett.* 11, 151 (1974).
**34** W.E. Caswell and F. Wilczek, *Phys. Lett.* 49B, 291 (1974).
**35** D.R.T. Jones, *Nucl. Phys.* B75, 531 (1974).
**36** W.E. Caswell, *Phys. Rev. Lett.* 33, 244 (1974).
**37** K.G. Chetyrkin, A.L. Kataev and F.V. Tkachov, *Phys. Lett.* 858, 277 (1979).
**38** M. Dine and J. Sapirstein, *Phys. Rev. Lett.* 43, 668 (1979).
**39** O.V. Tarasov, A.A. Vladamirov and A.Yu. Zharkov, *Phys. Lett.* 93B, 429 (1980).
**40** S. Ferrara and B. Zumino, *Nucl. Phys.* B79, 413 (1974).
**41** A. Salam and J. Strathdee, *Phys. Lett.* 51B, 353 (1974).
**42** D.R.T. Jones, *Nucl. Phys.* 87B, 127 (1975).
**43** D.R.T. Jones, *Phys. Lett.* 72B, 199 (1977).
**44** E.C. Poggio and H. Pendleton, *Phys. Lett.* 72B, 200 (1977).
**45** O.V. Tasarov, A.A. Vladamirov and A.Yu. Zharkov, *Phys. Lett.* 96B, 94 (1980).
**46** M. Grisaru, M. Rocek and W. Siegel, *Phys. Rev. Lett.* 45, 1063 (1980).
**47** S. Mandelstam, in *Proceedings of the 21st International Conference on High Energy Physics*, Paris, 1982, p. C3-331.
**48** P.S. Howe, K.S. Stelle and P.C. West, *Phys. Lett.* 124B, 55 (1983).
**49** A.J. Parkes and P.C. West, *Phys. Lett.* 138B, 99 (1984).
**50** D.R.T. Jones and L. Mezincescu, *Phys. Lett.* 138B, 293 (1984).
**51** A.J. Parkes and P.C. West, *Nucl. Phys.* B256, 340 (1985).
**52** D.R.T. Jones and A.J. Parkes, *Phys. Lett.* 160B, 267 (1985).
**53** S. Weinberg, *Phys. Rev. Lett.* 19, 1264 (1967).
**54** W. Buchmuller, Y.J. Ng and S.H.H. Tye, *Phys. Rev.* D24, 3003 (1981).
**55** W. Marciano, *Phys. Rev.* D20, 274 (1979).
**56** T. Goldman and D. Ross, *Nucl. Phys.* B171, 273 (1980).
**57** W. Marciano and A. Sirlin, *Phys. Rev. Lett.* 46, 163 (1981).
**58** S.L. Glashow and D.V. Nanopoulos, *Nature* 281, 464 (1979).
**59** J.D. Bjorken, *Phys. Rev.* 179, 1547 (1969).
**60** J.D. Bjorken and E.A. Paschos, *Phys. Rev.* 185, 1975 (1969).
**61** R.P. Feynman, *Phys. Rev. Lett.* 23, 1415 (1969).
**62** F.E. Close, *An Introduction to Quarks and Partons*, Academic Press, New York, 1979.
**63** C.G. Callan and D.J. Gross, *Phys. Rev. Lett.* 22, 156 (1969).
**64** P.C. Bosetti et al., *Nucl. Phys.* B142, 1 (1978).
**65** J.G.H. de Groot et al., *Phys. Lett.* 82B, 456 (1979).
**66** J.G.H. de Groot et al., *Z. Phys.* C1, 143 (1979).
**67** S.L. Adler, *Phys. Rev.* 143, 1144 (1966).
**68** D.J. Gross and C.H. Llewellyn Smith, *Nucl. Phys.* B14, 337 (1969).
**69** M.Y. Han and Y. Nambu, *Phys. Rev.* 139B, 1006 (1965).
**70** P.H. Frampton and J.J. Sakurai, *Phys. Rev.* D16, 572 (1977).
**71** R.A. Brandt and G. Preparata, *Nucl. Phys.* B27, 541 (1971).
**72** Y. Frishman, *Phys. Rev. Lett.* 25, 966 (1970).
**73** Y. Frishman, *Ann. Phys.* 66, 373 (1971).
**74** R. Jackiw, R. Van Royen and G.B. West, *Phys. Rev.* D2, 2473 (1970).
**75** J.G.H. de Groot et al., *Phys. Lett.* 82B, 292 (1979).
**76** H. Georgi and H.D. Politzer, *Phys. Rev.* D9, 416 (1974).
**77** G. Altarelli and G. Parisi, *Nucl. Phys.* B126, 298 (1977).
**78** B. DeWitt, *Phys. Rev.* 162, 1195 (1967).
**79** B. DeWitt, *Phys. Rev.* 162, 1239 (1967).
**80** J. Honerkamp, *Nucl. Phys.* B36, 130 (1971).
**81** J. Honerkamp, *Nucl. Phys.* B48, 269 (1972).
**82** G. 't Hooft, *Nucl. Phys.* B62, 444 (1973).
**83** R. Kallosh, *Nucl. Phys.* B78, 293 (1974).
**84** H. Kluberg-Stern and J.B. Zuber, *Phys. Rev.* D12, 467 (1975).
**85** H. Kluberg-Stern and J.B. Zuber, *Phys. Rev.* D12, 482 (1975).
**86** H. Kluberg-Stern and J.B. Zuber, *Phys. Rev.* D12, 3159 (1975).
**87** L.F. Abbott, *Nucl. Phys.* B185, 189 (1981).
**88** S. Weinberg, *Phys. Lett.* 91B, 51 (1980).
**89** R. Tarrach, *Nucl. Phys.* B196, 45 (1982).

**90** B.A. Ovrut and H.J. Schnitzer, *Phys. Lett.* 110B, 139 (1982).

**91** G. ’t Hooft and M. Veltman, *Ann. Inst. Henri Poincaré* 20, 69 (1974).

**92** S. Deser, J. Kay and K. Stelle, *Phys. Rev. Lett.* 38, 527 (1977).

# 6
# Quantum Chromodynamics

## 6.1
## Introduction

Here we treat quantum chromodynamics (hereafter QCD), the universally accepted theory of strong interactions. We discuss the limitations of perturbative techniques, especially the arbitrariness of the renormalization scheme. As an application we discuss the total annihilation cross section of electron and positron. Two and three jet processes in this annihilation present good agreement with perturbative calculations, at least at a one per mil level.

For nonperturbative calculations, we first treat instantons and show how they solve one problem, that of the fourth (or ninth) axial current and its missing Goldstone boson. Instantons create another problem of why strong interactions conserve CP to an accuracy of at least 1 in $10^9$.

Expansion in $1/N$, where $N$ is the generic number of colors, leads to successful qualitative explanations of phenomenological selection rules and of the absence of exotic mesons. Remarkably, baryons appear as topological solitons (Skyrmions) in chiral pion theories.

The most powerful quantitative method for handling nonperturbative QCD is on a space–time lattice which leads to automatic quark confinement at large distances. Monte Carlo simulations connect this strong coupling regime smoothly to the scaling regime at weak coupling and lead to quantitative estimates of the string tension, the interquark force, and glueball and hadron masses. Comparison with experiment works well within the (a few %) errors.

## 6.2
## Renormalization Schemes

QCD is the non-Abelian gauge theory based on an unbroken gauge group SU(3) with an octet of gluons as gauge bosons. Quarks are in the defining three-dimensional representation, and antiquarks in the conjugate representation. QCD is hence an extremely simple gauge field theory in the sense of specifying the classical Lagrangian density. The task of showing that QCD correctly describes the wealth of data on hadronic processes is, however, far from simple.

The QCD theory was first proposed in the early 1970s [1–4], although the idea of gauging color had been mentioned in 1966 by Nambu [5]. Key ingredients are, of course, the quarks, which were first suggested by Gell-Mann [6] and Zweig [7] (reprinted in Ref. [8]). Also, the color concept arose from the conflict of spin and statistics in the quark wavefunctions for the lowest-lying baryon states [9, 10]. But really it was the proof of renormalizability of Yang–Mills theory [10] and the establishment of the property of asymptotic freedom [11, 12] which led to the general acceptance of QCD.

Unlike the situation for quantum electrodynamics (QED) over 20 years earlier, there is no simple static property comparable to the Lamb shift [13] or anomalous moment [14, 15] in QED, so precise checks of QCD remain elusive beyond the 0,1% level. Our first task is to address the perturbation expansion of QCD and to illustrate the successes and limitations thereof.

Much of the initial preparation has been done by the discussion of dimensional regularization (Chapter 3) and of the renormalization group equations for QCD (Chapter 5). From the latter discussion, we could arrive at a running coupling constant $\alpha_s(Q^2)$ with

$$\alpha_s(Q^2) = \frac{12\pi}{(33 - 2N_f)\ln(Q^2/\Lambda^2)} \tag{6.1}$$

by using just the one-loop formula for the renormalization group $\beta$-function, that is, the first term of the series

$$\beta = -\frac{g^3}{48}\pi^2(33 - 2N_f) - \frac{g^5}{768\pi^4}(306 - 38N_f) + \cdots \tag{6.2}$$

We asserted in Chapter 5 that the scale $\Lambda$ in Eq. (6.1) must be determined phenomenologically. Actually, the evaluation of $\Lambda$ is quite complicated and always involves specification of the renormalization scheme, as we shall discuss; further, the phenomenological evaluation of $\Lambda$ necessitates theoretical computation up to at least two loops in all cases—one-loop comparison carries sufficient ambiguity to be meaningless.

Nevertheless, it makes good sense to write an “improved” perturbation expansion [16, 17] for QCD where the expansion parameter is taken to be the running coupling $\alpha_s(Q^2)$ rather than an energy-independent constant. This is more important in QCD than in QED because in the latter there is a natural choice ($Q^2 = 0$) for definition of the coupling constant. Also, $\alpha_5 \gg \alpha$, so the loop corrections in QCD are intrinsically much larger: only for large $Q^2$ is $\alpha_s(Q^2) \ll 1$ and perturbation theory reliable, but then we must use the $Q^2$-dependent $\alpha_s(Q^2)$.

The dependence of the expansion coefficients in the QCD improved perturbation series on the choice of renormalization scheme creates a situation with no well-defined answer because the series itself is not convergent. Nevertheless, one criterion in choosing a scheme is that of fastest apparent convergence (i.e., the scheme where the first one or two coefficients appear most rapidly decreasing). But let us illustrate some renormalization schemes explicitly for the case of the $e^+e^-$ annihilation, specifically for the ratio

$$R = \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} \tag{6.3}$$

The simplest renormalization scheme [18] is that of minimal subtraction (MS), which employs dimensional regularization of Feynman integrals and simply subtracts the divergence associated with a pole in the generic complex dimension plane. For example, if dimensional regularization gives the gamma function $\Gamma(2 - n/2)$, where $n$ is the generic dimension, we renormalize by subtracting the pole

$$\Gamma\left(2 - \frac{n}{2}\right) - \frac{2}{4-n} \tag{6.4}$$

An alternative to this [19] (see also Ref. [20]) is modified minimal subtraction ($\overline{\text{MS}}$), which subtracts from the finite part certain transcendental constants which are inevitably present in the dimensional regularization approach. Note that

$$\Gamma(z) = \frac{\Gamma(1+z)}{z} \tag{6.5}$$

$$= \frac{1}{z} + \Gamma'(1) + O(z) \tag{6.6}$$

$$= \frac{1}{z} - \gamma + O(z) \tag{6.7}$$

where $\gamma$ is the Euler–Mascheroni constant $\gamma = 0.5772$, defined by

$$\gamma = \lim_{p\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{p} - \ln p\right) \tag{6.8}$$

Also, the procedure of writing

$$\int \frac{d^n k}{(2\pi)^4} \tag{6.9}$$

means that one is arriving at a factor

$$\frac{\pi^2}{(2\pi)^4} = \frac{1}{(4\pi)^2} \tag{6.10}$$

which is really, in generic dimension, corrected by

$$(4\pi)^{(1/2)(n-4)} = 1 + \frac{1}{2}(n-4)\ln 4\pi + \cdots \tag{6.11}$$

and this means that a natural subtraction ($\overline{\text{MS}}$) to replace Eq. (6.4) (MS) is

$$\Gamma\left(2-\frac{n}{2}\right)-\frac{2}{4-n}+\gamma-\ln 4\pi \tag{6.12}$$

We shall see shortly how use of Eq. (6.12) ($\overline{\text{MS}}$) rather than Eq. (6.4) (MS) affects the apparent convergence of the QCD perturbation expansion of $R$ in Eq. (6.3). Changing from MS to $\overline{\text{MS}}$ also changes the value of the scale parameters $\Lambda_{\text{MS}}$ to $\Lambda_{\overline{\text{MS}}}$ in a way that we shall compute; this also changes the value of the coupling constant $\alpha_s(Q^2)$.

Before proceeding, let us note how dimensional regularization introduces an arbitrary scale ($\mu$) into the theory. When we cut off momentum integrals, this is obvious; using dimensional regularization it is slightly more subtle and arises because the coupling constant $g$ (which is dimensionless in $n = 4$) has dimension $M^{(4-n)/2}$, so we must redefine

$$g \to g\mu^{(4-n)/2} \tag{6.13}$$

if $g$ is to remain dimensionless. This is how the arbitrary mass scale $\mu$ enters in dimensional regularization.

Let us now proceed to discuss the calculation of the electron–positron annihilation cross-section. Consider $e^+(p_1) + e^-(p_2) \to X$, where $X$ has total momentum $q = (p_1 + p_2)$ and the squared center of mass energy is $s = q^2$ (Fig. 6.1). The relevant amplitude is

$$A\left(e^+e^- \to X\right) = \frac{2\pi e^2}{s}\bar{v}(p_1,\sigma)\gamma_\mu u(p_2,\sigma_2)\langle X|J_\mu|0\rangle \tag{6.14}$$

where $J_\mu$ is the electromagnetic current for quarks,

$$J_\mu \sum_f Q_f \bar{q}_f \gamma_\mu q_f \tag{6.15}$$

summed over the quark flavors. The $e^+e^-$ cross section is given by

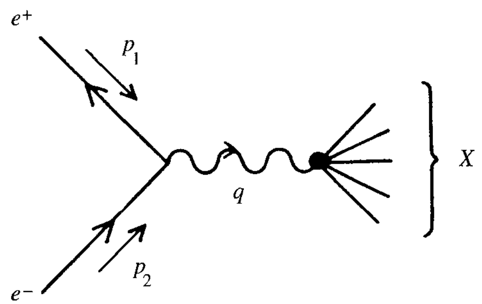



**Figure 6.1** Electron–positron annihilation.

$$\sigma = \sum_X \sigma\left(e^+e^- \to X\right) \tag{6.16}$$

$$= \frac{8\pi^2\alpha^2}{s^3} l_{\mu\nu} \sum_X (2\pi)^4 \delta^4(p_1 + p_2 - q)\langle X|J^\nu(0)|0\rangle \, \langle X|J^\mu(0)|0\rangle^* \tag{6.17}$$

where the leptonic piece is

$$l_{\mu\nu} = \frac{1}{4} \sum_{\sigma_1,\sigma_2} \bar{v}(p_1,\sigma_1)\gamma_\mu u(p_2,\sigma_2)\left[\bar{v}(p_1,\sigma_1)\gamma_\nu u(p_2,\sigma_2)\right]^* \tag{6.18}$$

$$= \frac{1}{4} \,\mathrm{Tr}(\not{p}_1\gamma_\mu \not{p}_2\gamma_\nu) \tag{6.19}$$

$$= \frac{1}{2}\left[q_\mu q_\nu - q^2 g_{\mu\nu} - (p_1 - p_2)_\mu (p_1 - p_2)_\nu\right] \tag{6.20}$$

where we have neglected the electron mass. The quark part of Eq. (6.17) can be related by unitarity to the imaginary part of the photon propagator; we may write

$$Q^{\mu\nu} = \sum_X (2\pi)^4 \delta^4(p_1 + p_2 - q)\langle 0|J_\mu(0)|X\rangle \, \langle 0|J_\nu(0)|X\rangle^* \tag{6.21}$$

$$= \int d^4x e^{iq\cdot x} \langle 0|\left[J_\mu(x), J_\nu(0)\right]|0\rangle \tag{6.22}$$

$$= 2\,\mathrm{Im}\,\pi^{\mu\nu} \tag{6.23}$$

where

$$\Pi^{\mu\nu}(q) = i \int d^4x e^{iq\cdot x} \langle 0|T\left(J_\mu(x) J_\nu(0)\right)|0\rangle \tag{6.24}$$

Conservation of $J_\mu$ dictates that

$$\Pi^{\mu\nu}(q) = \left(-g^{\mu\nu}q^2 + q^\mu q^\nu\right)\Pi(q) \tag{6.25}$$

In leading order ("zero loop") we may compute $\Pi^0(q)$ from Fig. 6.2 and find that

$$\mathrm{Im}\,\Pi^{(0)}(q) = \frac{1}{12\pi} 3 \sum_f Q_f^2 \tag{6.26}$$

The cross section is therefore

$$\sigma = \frac{4\pi\alpha^2}{3s} 3 \sum_f Q_f^2 \tag{6.27}$$

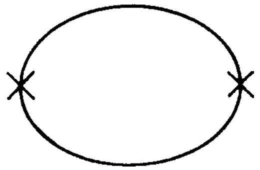


**Figure 6.2** Vacuum polarization.

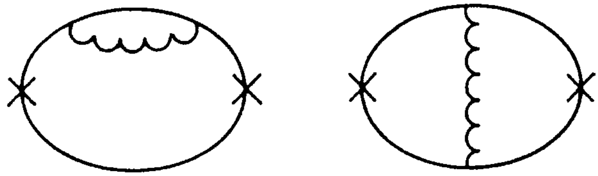


**Figure 6.3** Radiative corrections.

The process $e^+e^- \to \mu^+\mu^-$ has cross section

$$\sigma\left(e^+e^- \to \mu^+\mu^-\right) = \frac{4\pi\alpha^2}{3s} \tag{6.28}$$

and hence

$$R = 3\sum_f Q_f^2\left[1 + a_1\frac{\alpha_s}{\pi} + a_2\left(\frac{\alpha_s}{\pi}\right)^2 + \cdots\right] \tag{6.29}$$

At the next order we must calculate the diagrams of Fig. 6.3, and this gives [16, 17]

$$a_1 = 1 \tag{6.30}$$

by a calculation similar to that in QED [21]. The two-loop calculation is less trivial and involves the diagrams of Fig. 6.4; here the dependence on renormalization scheme becomes obvious. The result is [22, 23]

$$\begin{aligned} a_2(\text{MS}) &= a_2(\overline{\text{MS}}) + (\ln 4\pi - \gamma)\frac{33 - 2n_f}{12} \\ &= 7.35 - 0.44n_f \end{aligned} \tag{6.31}$$

$$\begin{aligned} a_2(\overline{\text{MS}}) &= \left[\frac{2}{3}\zeta(3) - \frac{11}{12}\right]n_f + \frac{365}{24} - 11\zeta(3) \\ &= 1.98 - 0.11n_f \end{aligned} \tag{6.32}$$

where $n_f$ is the number of quark flavors. Setting $n_f = 4$, for example, gives $a_2(\text{MS}) = +5.6$ and $a_2(\overline{\text{MS}}) = +1.52$, so that the MS scheme gives greater apparent convergence in Eq. (6.29) that does the MS scheme.

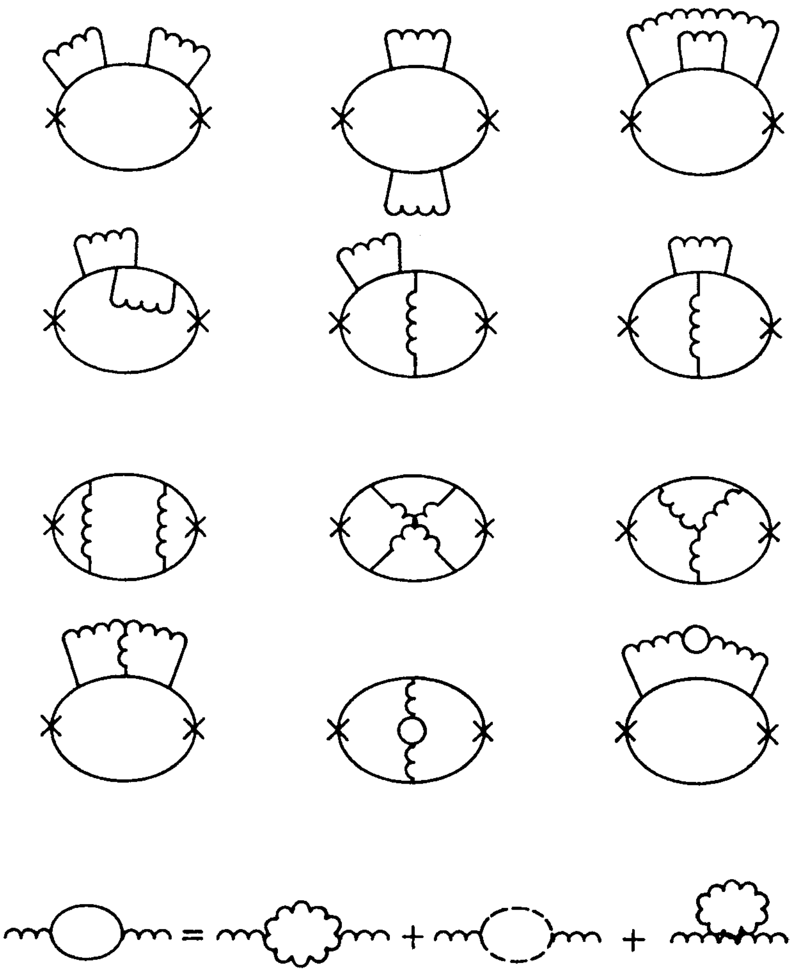


**Figure 6.4** Two-loop corrections.

We can compute the corresponding changes in the scale $\Lambda$ and the coupling strength $\alpha$. The renormalization group $\beta$ function to two loops is

$$\beta = \frac{g^3}{48\pi^3}(33 - 2n_f) - \frac{g^5}{384\pi^4}(153 - 19n_f) \tag{6.33}$$

and solving $dg/d\ln\mu = -\beta$ gives a formula

$$\alpha(Q^2) = \frac{12\pi}{(33 - 2n_f)\ln(q^2/\Lambda^2)} \cdot \left[1 - 6\frac{165 - 19n_f}{(33 - 2n_f)^2}\frac{\ln\ln(Q^2/\Lambda^2)}{\ln(Q^2/\Lambda^2)}\right] \tag{6.34}$$

Since $R$ in Eq. (6.29) is a physical quantity, it is obvious that $\Lambda_{\text{MS}} \neq \alpha_{\overline{\text{MS}}}$. In fact, one finds immediately that

$$\alpha_{\text{MS}} = \alpha_{\overline{\text{MS}}} - \frac{12\pi(\ln 4\pi - \gamma)}{(33 - 2n_f)[\ln(q^2/\Lambda^2)]^2} \tag{6.35}$$

This change can be absorbed into the scale $\Lambda$ by putting

$$\Lambda_{\overline{\text{MS}}} = \Lambda_{\text{MS}} e^{(1/2)(\ln 4\pi - \gamma)} \tag{6.36}$$

Here the multiplicative factor is about 2.6, so the value of the QCD scale, like $\alpha_s(Q^2)$, depends significantly on the renormalization scheme. Also, it is clear that a two-loop calculation is necessary to distinguish between schemes: At one-loop order the different $\Lambda$ and $\alpha_s$ may be interchanged, the error being only of higher order.

The experimental situation for $R$ is shown in Fig. 6.5. The agreement with QCD is impressive, although the experimental errors are too large to allow confirmation of the higher-order corrections.

There is an infinite number of possible renormalization schemes, in addition to MS and $\overline{\text{MS}}$. A third example is the momentum subtraction scheme (MOM) [24–27]. Here the three-gluon vertex is defined to be $g$ when all three gluons satisfy $p^2 = \mu^2$. Working in Landau gauge gives in Eq. (6.29) [26]

$$a_2(\text{MOM}) = 0.74n_f - 4.64 \tag{6.37}$$

which gives, for $n_f = 4$, $a_2(\text{MOM}) = -1.68$, so that the apparent convergence is no better than for the $\overline{\text{MS}}$ scheme. The MOM scheme is more similar to the

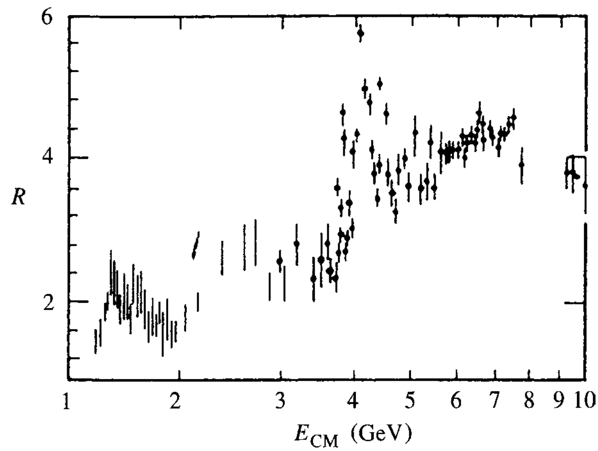



**Figure 6.5** Experimental data on $\sigma(e^+e^- \to \text{hadrons})/\sigma(e^+e^- \to \mu^+\mu^-)$. [After Particle Data Group, Rev. Mod. Phys. 56, S64 (1984).]

procedure normally used in QED but is gauge dependent and hence somewhat less convenient than $\overline{\text{MS}}$.

Although we have chosen to focus on $e^+e^-$ annihilation, the choice of renormalization scheme similarly affects the perturbation expansion for all physical processes. (For a review, see, for example, Ref. [28].)

One may ask whether there is some principle to select an optimal renormalization scheme? One suggestion [29] which has aroused interest [30–37] is that one should choose a scheme which is least sensitive to the choice of scheme. The argument is that the sum of the perturbative series is independent of the scheme, and hence the best approximations should also be.

This suggestion assumes (unjustifiably, see below) that the perturbation expansion is summable. In QED, the radius of convergence of the perturbation expansion of Green's functions

$$G(\alpha) = \sum_n a_n \alpha^n \tag{6.38}$$

is known to be zero by Dyson's argument [38], which points out that for all $\alpha < 0$, the vacuum would be unstable under electron–positron pair creation. This is because a virtual pair created with sufficiently small separation may become real by separating to larger distance. Thus the radius of convergence of Eq. (6.38) vanishes.

In QCD we may not apply Dyson's argument directly. Instead, we may usefully construct

$$G(\alpha) = \sum_n \frac{a_n}{n!} \int_0^\infty e^{-t} (\alpha t)^n \, dt \tag{6.39}$$

$$= \int_0^\infty e^{-t} f(t\alpha) \, dt \tag{6.40}$$

where, in Eq. (6.39), we used the integral representation of $\Gamma(n+1) = n!$. In Eq. (6.40), $f(x)$ is the Borel transform with expansion

$$f(x) = \sum_n \frac{a_n}{n!} x^n \tag{6.41}$$

The radius of convergence of the series for $G(\alpha)$ in Eq. (6.38) is given by $R$:

$$\frac{1}{R} = \lim_{n\to\infty} \sup\big(|a_n|^{1/n}\big) \tag{6.42}$$

while that for $f(x)$ is $R$:

$$\frac{1}{R'} = \lim_{n\to\infty} \sup\left|\frac{a_n}{n!}\right|^{1/n} \tag{6.43}$$

From Eqs. (6.42) and (6.43) it follows that for $R > 0$ then $R' = \infty$. Conversely, any singularity in the Borel plane implies that $R = 0$ (see, e.g., Ref. [39]).

In a quantum field theory, the Borel analysis can be related to the existence of finite action solutions to the classical field equations, as shown first by Lipatov [40] and expounded by others [41–45]. Consider a Green's function [here $\phi(x)$ is a generic field of arbitrary spin, etc.]

$$\langle 0|T\left(\phi(x_1)\cdots\phi(x_n)\right)|0\rangle = \frac{\int[D\phi]e^{-S[\phi]}\phi(x_1)\cdots\phi(x_n)}{\int[D\phi]e^{-S[\phi]}} \tag{6.44}$$

We may scale out the coupling strength $(\alpha)$ and rewrite the numerator $N(\alpha)$ by making use of

$$\alpha\int_0^\infty dt\,\delta\big(\alpha t - S[\phi]\big) = 1 \tag{6.45}$$

in the form

$$N(\alpha) = \alpha\int_0^\infty dt F(\alpha t)e^{-t} \tag{6.46}$$

where the Borel transform $F(z)$ is

$$F(z) = \int[D\phi]\delta\big(z - S[z]\big)\phi(x_1)\cdots\phi(x_n) \tag{6.47}$$

Where are the singularities, if any, of Eq. (6.47)? Consider the similar integral

$$\int dy_1\,dy_2\cdots dy_n\delta\big(z - f(y_1, y_2\cdots y_n)\big) = \int_\sigma d\sigma|\nabla f|^{-1} \tag{6.48}$$

where the surface $\sigma$ is where $f(y) = z$. The result diverges if anywhere on $\sigma$ there is a point at which

$$|\nabla f|^2 = \sum_i\left|\frac{\partial f}{\partial y_i}\right|^2 = 0 \tag{6.49}$$

which implies that $\partial f/\partial y_i = 0$ for all $i$. Similarly Eq. (6.47) is singular if for some $S[\phi] = z$ there is a point satisfying

$$\frac{\partial S[\phi]}{\partial\phi(x)} = 0 \tag{6.50}$$

for all $x$. But Eq. (6.50) is just the classical field equation in Euclidean space. Thus, if there is a finite action classical solution, it renders the Borel transform singular and implies that the original perturbation expansion converges nowhere. QCD possesses finite action classical solutions, and hence its perturbation expansion has zero radius of convergence.

From this result it emerges that there is no unique way to choose a renormalization scheme. One must be more pragmatic (as, e.g., Llewellyn Smith [28] is) and check the apparent convergence physical process by physical process.

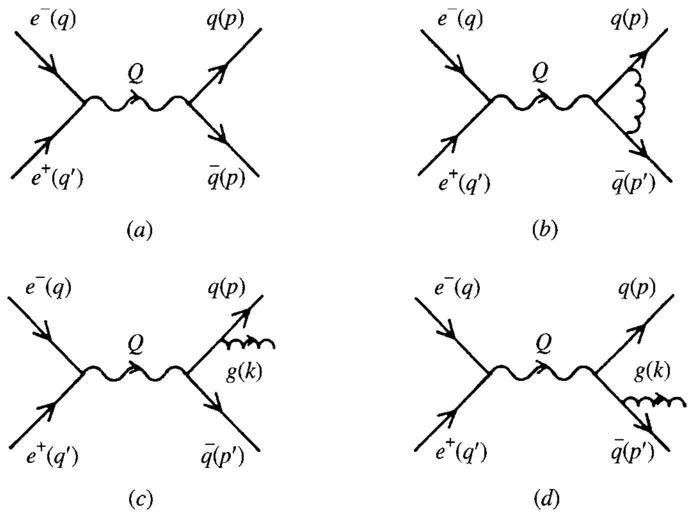



**Figure 6.6** Three-jet diagrams.

## 6.3
## Jets in Electron–Positron Annihilation

We have already discussed the total electron–positron annihilation cross section, and now we wish to examine the possible final states in more detail. For example, to zeroth orders in QCD we may consider Fig. 6.6*a*, where a quark–antiquark pair is produced. Assuming that these quarks somehow change to hadrons, we may naively expect [46–48] from a parton model picture [49, 50] that the final state will contain back-to-back two-jet structures. This was strikingly confirmed experimentally at SLAC in 1975 [51] (see also Refs. [52] and [53]), with an angular distribution (see below) confirming the spin-$\frac{1}{2}$ nature of the quarks produced.

In a perturbation expansion we encounter at next order the diagrams depicted in Fig. 6.6*b* through *d*. These diagrams exhibit two types of singular behavior: infrared divergences associated with the masslessness of the gluon, and mass singularities associated with the masslessness of the final-state quarks. Such singularities are familiar in quantum electrodynamics [54–57] and necessitate suitable averaging over accessible states (see especially Kinoshita [56], and Lee and Nauenberg [57]). In the present case we must take care in combining Fig. 6.6*b* with Fig. 6.6*c* and *d* because when one of the three final particles is near the kinematic boundary—either with very low energy or nearly collinear with one of the other two particles—the event is really a two-jet, rather than a three-jet, configuration. The theory reflects this physical fact and yields sensible nonsingular answers only when a correct averaging is made.

Three-jet events were first discussed by Ellis, Gaillard, and Ross [58] and by DeGrand, Ng, and Tye [59]. The handling of mass singularities was explained in the work of Sterman and Weinberg [60]; there are several further theoretical analyses [61–66]. Three jet events were discovered experimentally (see below) at DESY in 1979 [67–70], and provided compelling evidence for the physical existence of the gluon.

Our present discussion follows most closely that of Ref. [71], to which we refer for further details. Let the four-momenta be defined by $e^-(q) + e^+(q') \to q(p) + \bar{q}(p') + g(k)$. Let $Q_\mu = (q + q')_\mu = (p + p' + k)_\mu$ and $s = Q^2$. The zeroth-order process corresponding to Fig. 6.6*a* gives the squared matrix element

$$|m|^2 = \frac{1}{s^2} l_{\mu\nu} H_{\mu\nu} \tag{6.51}$$

where for massless unpolarized leptons

$$l_{\mu\nu} = \frac{1}{4} \sum_{S,S'} \langle 0|j_\mu|e^+e^-\rangle \, \langle 0|j_\nu|e^+e^-\rangle^* \tag{6.52}$$

$$= \frac{e^2}{4E^2}(q, q')_{\mu\nu} \tag{6.53}$$

in which the shorthand notation is

$$(A, B)_{\mu\nu} \equiv (A_\mu B_\nu + A_\nu B_\mu - g_{\mu\nu} A \cdot B) \tag{6.54}$$

This follows because

$$l_{\mu\nu} = \frac{1}{4} \mathrm{Tr}\left( \gamma_\mu \frac{\not{q}}{2E} \gamma_\mu \frac{\not{q}'}{2E} \right) \tag{6.55}$$

Similarly,

$$H_{\mu\nu} = \frac{e_f^2}{E^2}(p, p')_{\mu\nu} \tag{6.56}$$

where $e_f$ is the electric charge for the quark flavor $f$. Using

$$(A, B)_{\mu\nu}(C, D)_{\mu\nu} = 2\big[(A \cdot C)(B \cdot D) + (A \cdot D)(B \cdot C)\big] \tag{6.57}$$

gives

$$|m|^2 = \frac{e^2 e_f^2}{4E^4}\left(1 + \cos^2\theta\right) \tag{6.58}$$

where $\theta$ is the center-of-mass scattering angle. Hence

$$d\sigma = \pi \frac{1}{8\pi^3} \int d^3p \, d^3p' |m|^2 \delta^4(Q - p - p') \tag{6.59}$$

$$\left(\frac{d\sigma}{d\Omega}\right)_0 = \frac{\alpha^2}{4S}\left(1 + \cos^2\theta\right) \sum_{f,c} \left(\frac{e_f}{e}\right)^2 \tag{6.60}$$

$$\sigma_0 = \frac{4\pi\alpha^2}{3S} \sum_{f,c} \left(\frac{e_f}{e}\right)^2 \tag{6.61}$$

These are the results for the zeroth-order cross sections.

Now consider Fig. 6.6*c* and *d*. We shall calculate the three-jet cross section at lowest nontrivial order in $\alpha_s$, then obtain the two-jet cross section by subtraction from the known total cross section; at order $\alpha_s$ we may clearly not produce more than three jets in the final state. Care will be necessary to separate those pieces of Fig. 6.6*c* and *d* which are so near the kinematic boundary that they will appear as two-jet events.

For these new diagrams, $l_{\mu\nu}$ is as in Eq. (6.53), but now we have

$$H_{\mu\nu} = \sum_{\substack{s,t',t\\ c,f}} \langle q\bar{q}g|J_\mu|0\rangle \, \langle q\bar{q}g|J_\nu|0\rangle^* \tag{6.62}$$

where $t_\mu$ is the polarization vector of the gluon. The relevant color summation is then

$$\mathrm{Tr}\big(A_\mu^a \lambda^a A_\nu^b \lambda^b\big) = \frac{1}{2k_0} t_\mu t_\nu \delta^{ab}\, \mathrm{Tr}\big(\lambda^a \lambda^b\big) \tag{6.63}$$

$$= \frac{8}{k_0} t_\mu t_\nu \tag{6.64}$$

Using this, one finds for the sum of the two diagrams, the Dirac trace

$$H_{\mu\nu} = \frac{1}{8p_0 p_0' k_0} \sum_{f,t} e_f^2 g^2\, \mathrm{Tr}\left\{\not{p}\left[\not{t}\frac{1}{p\cdot k}(\not{p}+\not{k})\gamma_\mu - \gamma_\mu \frac{1}{p'\cdot k}(\not{p}'+\not{k})\not{t}\right]\right.$$

$$\left. \cdot \left[\gamma_\nu \frac{1}{p\cdot k}(\not{p}+\not{k})\not{t} - \not{t}\frac{1}{p'\cdot k}(\not{p}'+\not{k})\gamma_\nu\right]\right\} \tag{6.65}$$

In the notation of Eq. (6.54), this becomes, using $\sum_t t_\mu t_\nu = -g_{\mu\nu}$

$$H_{\mu\nu} = \frac{2}{p_0 p_0' k_0} \sum_f e_f^2 g^2 \left\{ \frac{p\cdot p'}{(p\cdot k)(p'\cdot k)}\big[2(p, p')_{\mu\nu} + (k, p+p')_{\mu\nu}\big]\right.$$

$$+ \frac{1}{p\cdot k}\big[(p', p+k)_{\mu\nu} - (p, p)_{\mu\nu}\big]$$

$$\left. + \frac{1}{p'\cdot k}\big[(p, p'+k)_{\mu\nu} - (p', p')_{\mu\nu}\big]\right\} \tag{6.66}$$

The squared matrix element is given by

$$|m|^2 = \left(\frac{1}{4E^2}\right)^2 l_{\mu\nu} H_{\mu\nu} \tag{6.67}$$

where $Q^2 = (2E)^2$ is the center-of-mass energy. Using Eq. (6.57) and identities such as $(Q \cdot q) = (Q \cdot q')$, one obtains

$$\begin{aligned}|m|^2 = \frac{1}{16E^6 p_0 p_0' k_0} \sum_f (ee_f g)^2 \Bigg\{ &\frac{p \cdot p'}{(p \cdot k)(p' \cdot k)} \big[2(p \cdot q)(p' \cdot q') \\ &+ 2(p \cdot q')(p' \cdot q) - 2(k \cdot q)(k \cdot q') + (Q \cdot q)(Q \cdot k)\big] \\ &+ \frac{1}{p \cdot k}\big[(Q \cdot q)(Q \cdot p') - 2(p \cdot q)(p \cdot q') - 2(p' \cdot q)(p' \cdot q')\big] \\ &+ \frac{1}{p' \cdot k}\big[(Q \cdot q)(Q \cdot p) - 2(p \cdot q)(p \cdot q') - 2(p' \cdot q)(p' \cdot q')\big]\Bigg\}\end{aligned} \tag{6.68}$$

Let us take again the center-of-mass system in which

$$q = (E, \mathbf{q}); \qquad q' = (E, -\mathbf{q}) \tag{6.69}$$

$$p = (p_0, \mathbf{p}), \qquad p' = (p_0', \mathbf{p}'), \qquad k = (k_0, \mathbf{k}) \tag{6.70}$$

and define the variables $x$, $y$, and $z$ by

$$x = \frac{p_0}{E}, \qquad y = \frac{p_0'}{E}, \qquad z = \frac{k_0}{E} \tag{6.71}$$

satisfying

$$x + y + z = 2 \tag{6.72}$$

The differential cross section for the process $e^-(q) + e^+(q') \to q(p) + \bar{q}(p') + g(k)$ is given by

$$d\sigma = \pi \left(\frac{1}{8\pi^3}\right)^2 \int |m|^2 d^3p\, d^3p'\, d^3k \delta^4(p + p' + k - Q) \tag{6.73}$$

and it can be shown by straightforward manipulations that

$$d\sigma = \frac{2}{3E^2}\alpha^2 \frac{g^2}{4\pi}\, dx\, dy \frac{x^2 + y^2}{(1-x)(1-y)} \sum_f \left(\frac{e_f}{e}\right)^2 \tag{6.74}$$

When we consider a jet cross section, we must acknowledge that we shall not know which of the jets arises from $q$, $\bar{q}$, and $g$. Let the jets be numbered 1, 2, and 3 and let us define more symmetrically the variables

$$x_1 = \frac{p_{10}}{E}, \qquad x_2 = \frac{p_{20}}{E}, \qquad x_3 = \frac{p_{30}}{E} \tag{6.75}$$

where $p_{i\mu}$ is the momentum of the $i$th jet. Since $x_{+}x_2 + x_3 = 2$, we may plot each three-jet event (six times) within a triangular Dalitz plot: an equilateral triangle of perpendicular height (Fig. 6.7) equal to 2. All points actually lie in the inscribed triangle $ABC$. We may restrict attention to the region $x_1 \leqslant x_2 \leqslant x_3$ which is the triangle $AOD$, also shown in Fig. 6.7$b$. The three-jet cross section is now

$$d\sigma_{3\text{jet}} = \frac{4}{3E^3}\alpha^2 \frac{g^2}{4\pi}\, dx_1\, dx_2 \sum_f \left(\frac{e_f}{e}\right)$$
$$\cdot \left[\frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)} + \frac{x_2^2 + x_3^2}{(1-x_2)(1-x_3)} + \frac{x_3^2 + x_1^2}{(1-x_3)(1-x_1)}\right] \tag{6.76}$$

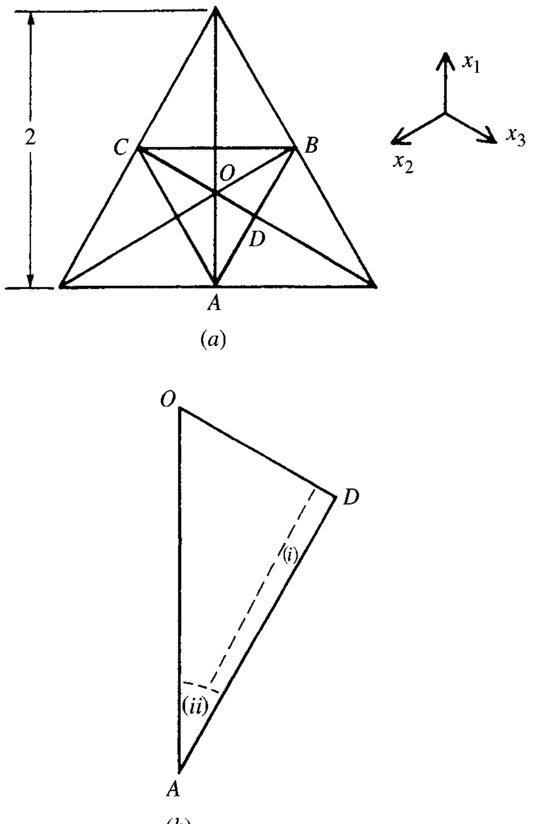



**Figure 6.7** Dalitz plot.

On the line $AD$, $x_3 = 1$ and the second and third terms in Eq. (6.76) become singular. Regions (i) and (ii) of Fig. 6.7*b* reflect these kinematic boundary regions for two situations in which the event will appear as only two jets rather than three jets. These are:

(i) Two of the final momenta are within small angle $\leqslant 2\delta$.
(ii) One of the final particles has energy less than the small value $2\epsilon E$, where $\epsilon \ll 1$, and the angle between the other two particles is between $(\pi - \delta)$ and $\pi$.

To obtain the three-jet cross section we must integrate Eq. (6.76) over the Dalitz plot, excluding regions (i) and (ii). To establish these excluded regions, consider the three three-momenta of the jets $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ adding to zero in the center of mass and width $|\mathbf{p}_3| \geqslant |\mathbf{p}_2| \gg |\mathbf{p}_1|$. The angles between their directions satisfy $\theta_{12} + \theta_{23} + \theta_{31} = \pi$. From the cosine rule

$$x_3^2 = x_1^2 + x_2^2 + 2x_1x_2 \cos\theta_{12} \tag{6.77}$$

and $x_1 + x_2 + x_3 = 2$, one finds that

$$\sin\frac{\theta_{12}}{2} = \sqrt{\frac{1 - x_3}{x_1x_2}} \tag{6.78}$$

and cyclic permutations thereof.

In region (ii) we need to have $x_1 \leqslant 2\epsilon$ and $0 \leqslant (\pi - \theta_{23}) \leqslant \delta$. Putting $1 - x_3 = \zeta \ll 1$ and assuming that $2\epsilon < \delta$, this boundary corresponds to $\zeta(x_1 - \zeta) \leqslant \delta^2/4$.

In region (i) we require simply $\theta_{12} \leqslant 2\delta$, and since Eq. (6.76) gives

$$\sin\frac{1}{2}\theta_{12} \simeq \sqrt{\frac{\zeta}{x_1(1 - x_1)}} \tag{6.79}$$

this corresponds to $\zeta \leqslant x_1(1 - x_1)\delta^2$.

Integration of the differential cross-section equation (6.74) over the Dalitz plot excluding these two-jet regions eventually gives

$$\sigma(3\text{-jet}) = \frac{1}{e^2}\alpha\frac{g^2}{3\pi}\left[(3 + 4\ln 2\epsilon)\ln\delta + \frac{4\pi^2 - 21}{12}\right]\sum_f\left(\frac{e_f}{e}\right)^2 \tag{6.80}$$

Now at this order $O(g^2)$, we know that

$$\delta(\text{total}) = \sigma(2\text{-jet}) + \sigma(3\text{-jet}) \tag{6.81}$$

In Section 6.2 we have seen that

$$\sigma(\text{total}) = \frac{\pi\alpha^2}{E^2}\sum_f\left(\frac{e_f}{e}\right)^2\left(1 + \frac{g^2}{4\pi^2} + \cdots\right) \tag{6.82}$$

and hence by subtraction we arrive at the formula [60]

$$\frac{d\sigma}{d\Omega}(\text{2-jet}) = \left(\frac{d\sigma}{d\Omega}\right)_0 \left\{1 - \frac{g^2}{3\pi^2}\left[(3 + 4\ln\epsilon)\ln\delta + \frac{2\pi^2 - 15}{6}\right]\right\} \tag{6.83}$$

with $(d\sigma/d\Omega)_0$ given by Eq. (6.60). Of course, the important theoretical point is that the infrared singularities associated with the massless gluon, and the mass singularities associated with massless quarks, have canceled in the final expressions (6.80) and (6.83) because we have correctly averaged the final states.

From Eq. (6.83) we see that the angular distribution $(1 + \cos^2\theta)$ of Eq. (6.60) holds also for the two-jet events. The experimental verification of this at SLAC [51] gave strong support for the spin-half nature of quarks. The jettiness of events was measured by determining the sphericity $S$ defined by [51, 58, 72]

$$S = \frac{3(\sum_i p_{\perp i}^2)_{\min}}{2\sum_i p_i^2} \tag{6.84}$$

where the sum is over final hadrons $i$ with three momenta of magnitude $P_I$. The jet axis is varied so that the sum of transverse squared momenta, relative to the chosen axis, is minimized. For an isotropic event, $S = 1$, whereas for perfect jets, $S = 0$.

It turns out that, theoretically, $S$ is infrared divergent and a more advantageous criterion for jettiness is to use spherocity $S'$, defined by [73]

$$S' = \left(\frac{4}{\pi}\right)^2 \left(\frac{\sum_i |p_{\perp i}}{\sum_i |p_i|}\right)^2_{\min} \tag{6.85}$$

This has the same range of values $1 \geqslant S' > 0$ as $S$ but is free of mass singularities. Another possibility is the thrust $T$, defined by [74]

$$T = \frac{2(\sum_i' p_{\parallel i})_{\max}}{\sum_i |p_i|} \tag{6.86}$$

where the numerator sum is over all particles in one hemisphere. This is similarly suitable for perturbative calculations; it varies between $T = \frac{1}{2}$(isotropy) and $T = 1$(perfect jets). Thrust is often employed by experimentalists: Fig. 6.8 displays a typical three-jet result at PETRA. As already mentioned, the clear observation of such three-jet final states supports the QCD analysis; the dashed line in Fig. 6.8 is from the quark–antiquark–gluon model we have been discussing.

## 6.4 Instantons

We have already mentioned how the existence of finite action classical solutions leads to zero radius of convergence of the QCD perturbation expansion.

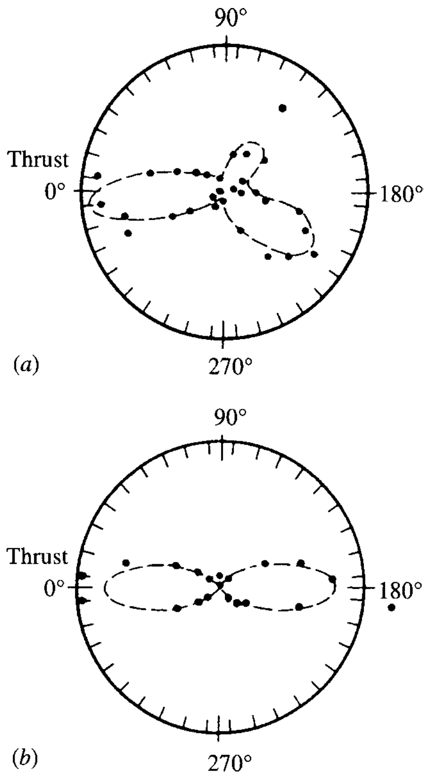



**Figure 6.8** Examples of jet data compared to theoretical QCD prediction (dashed line). [After D.P. Barber et al., Phys. Rev. Lett. 43, 830 (1979).]

Here we consider the instanton classical solution [75] and discuss how it solves one important problem, the U(1) problem, yet creates another serious difficulty, the strong CP problem.

Let us first state the U(1) problem. The Lagrangian for QCD with $n_f$ flavors is

$$L_{\text{QCD}} = -\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} + \sum_{k=1}^{n_f} q_k(i\not{D} - M_k)q_k \tag{6.87}$$

where $k$ runs over $k = u, d, s, \ldots$. Hadronic symmetries such as electric charge, Heisenberg isospin, and flavor SU(3) act only on this flavor label since hadrons are all color singlets. If we consider just two flavors, $n_f = 2$ and $k = u, d$, then in the limit of massless quarks $L_{\text{QCD}}$ possesses the chiral symmetry $\text{SU}(2) \times \text{SU}(2)$, since regarding $(u, d)$ as an SU(2) doublet we may make separate rotations of $q_{kL}$ and $q_{kR}$ in

$$\bar{q}_k\not{D}q_k \equiv \bar{q}_{kL}\not{D}q_{kL} + \bar{q}_{kR}\not{D}q_{kR} \tag{6.88}$$

This $\text{SU}(2) \times \text{SU}(2)$ is close to being an exact symmetry of the strong interactions. The diagonal SU(2) isospin subgroup is a symmetry broken by $M_u \neq M_d$, but since $M_u$, $M_d$, $(M_d - M_u) \ll \Lambda_{\text{QCD}}$, it is manifest as an almost exact symmetry of

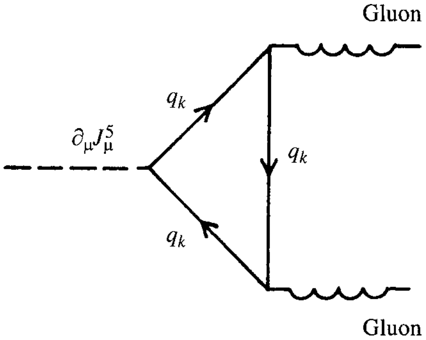



**Figure 6.9** Color triangle anomaly in $\partial_\mu J^5_\mu$.

strong interactions. The remainder of the group is a Nambu–Goldstone symmetry with three massless Goldstone bosons, the pseudoscalar pions $\pi^\pm$ and $\pi^0$. The smallness of the pion mass compared to all other hadron masses reveals how good a symmetry $SU(2) \times SU(2)$ is. It underlies the success of current algebra and PCAC computations in the 1960s.

The $L_{QCD}$ has two other global U(1) symmetries. One is the vectorial one associated with conservation of

$$J_\mu = \sum_k \bar{q}_k \gamma_\mu q_k \tag{6.89}$$

which is simply the baryon number which is exactly conserved in QCD. The second (problematic) one is associated with the current

$$J^5_\mu \sum_k \bar{q}_k \gamma_\mu \gamma_5 q_k \tag{6.90}$$

corresponding to the chiral transformation

$$q_k \to e^{i\alpha\gamma_5} q_k \tag{6.91}$$

This U(1) symmetry must either be a manifest symmetry, in which case all massive hadrons should be parity-doubled—clearly not true in nature—or it is broken spontaneously, leading to a fourth pseudoscalar Goldstone boson, $\eta$. Then, why is the $\eta$ at about 550 MeV when the pion is only 140 MeV? Weinberg [76] showed that one would expect $M_\eta < \sqrt{3} M_\pi$ from standard estimates of Goldstone boson masses. One can generalize to $n_f = 3$ and $SU(3) \times SU(3)$ and make similar arguments, but this larger symmetry is more badly broken and it is more perspicuous to focus on the excellent $SU(2) \times SU(2)$ symmetry. The "U(1) problem," then, is that the $\eta$ mass is too much larger than the $\pi$ mass.

The first step to solution of the U(1) problem is to observe that there is a color anomaly in $\partial_\mu J^5_\mu$ from the triangle diagram of Fig. 6.9. The standard calculation (see Chapter 3) gives

$$\partial_\mu J_\mu^5 = \frac{n_f g^2}{8\pi^2} \operatorname{Tr}(F_{\mu\nu}\mathscr{F}_{\mu\nu}) \tag{6.92}$$

where

$$\mathscr{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F_{\alpha\beta} \tag{6.93}$$

But we may use

$$\operatorname{Tr}(F_{\mu\nu}\mathscr{F}_{\mu\nu}) = \partial_\mu K_\mu \tag{6.94}$$

$$K_\mu = 2\epsilon_{\mu\nu\alpha\beta} \operatorname{Tr}\left( A_\nu \partial_\alpha A_\beta - \frac{2}{3} i g A_\nu A_\alpha A_\beta \right) \tag{6.95}$$

[we are using matrix notation $F_{\mu\nu} = F^a_{\mu\nu}T^a$ and $A_\mu = A^a_\mu T^a$, where $\operatorname{Tr}(T^a T^b) = \frac{1}{2}\delta^{ab}$, e.g., $T^a = \frac{1}{2}\lambda^a$ for Gell-Mann matrices] to define a gauge-dependent current

$$\mathscr{J}^5_\mu = J^5_\mu - \frac{n_f g^2}{8\pi^2} K_\mu \tag{6.96}$$

which remains conserved. We may write

$$\dot{Q}^5 = \int d^3x\, \partial_0 J^5_0 \alpha \int_S \mathbf{K} \cdot \mathbf{dS} \tag{6.97}$$

and provided that **K** falls off sufficiently rapidly for the background fields, $Q^5$ is still conserved and the U(1) problem persists. The anomaly alone is insufficient.

The U(1) problem was first resolved by 't Hooft [77, 78], who considered the change in $Q^5$ between the distant past and distant future according to

$$\Delta Q^5 = \int dx_0 \dot{Q}^5 = \int d^4x\, \partial_\mu J^5_\mu \tag{6.98}$$

$$= \frac{g^2 n_f}{8\pi^2} \int d^4x \operatorname{Tr}(F_{\mu\nu}\mathscr{F}_{\mu\nu}) \tag{6.99}$$

The key point is that there exist pure gauge configurations—instantons—such that $\Delta Q^5$ is nonzero. Hence, $Q^5$ is not a conserved charge and the extra Goldstone state is eliminated; in other words, because of instantons, the $\eta$ mass is not constrained by Goldstone boson mass formula to be less than $\sqrt{3}$ times the pion mass, after all.

To exhibit the instanton configuration, it is sufficient to study a pure SU(2) Yang–Mills theory. The reason is that the considerations of topological mappings for SU(2) generalize with no essential change to the SU(3) of QCD. The instantons are classical solutions in Euclidean space, but correspond to a quantum mechanical tunneling phenomenon in Minkowski space.

Topologically, as we shall explain, the group SU(2) is the three-sphere $S_3$; that is, a three-dimensional sphere where the surface of the Earth is a two-sphere $S_2$. From Eq. (6.97) we are interested in the mappings of the group on another three-sphere $S_3$ at infinity in Euclidean four-space. The relevant homotopy group is

$$\Pi_3\big(\mathrm{SU}(2)\big) = Z \tag{6.100}$$

where $Z$ is the set of all integers. There is a topological charge, or Pontryagin index, $\nu$, defined by

$$\nu = \frac{g^2}{16\pi^2}\int \mathrm{Tr}(F_{\mu\nu}\mathscr{F}_{\mu\nu})\, d^4x \tag{6.101}$$

so that in Eq. (6.97),

$$\Delta Q^5 = 2n_f \nu \tag{6.102}$$

Why is SU(2) topologically the same as $S_3$? The general SU(2) matrix may be written

$$g = a + i\mathbf{b}\cdot\boldsymbol{\sigma} \tag{6.103}$$

with

$$a^2 + |\mathbf{b}|^2 = 1 \tag{6.104}$$

which defines an $S_3$. At infinity we want $F_{\mu\nu}$ to vanish asymptotically, so $A_\mu$ must become a gauge transform of $A_\mu = 0$, that is,

$$A_\mu(x) \xrightarrow{r\to\infty} -\frac{i}{g}(\partial_\mu\Omega)\Omega^{-1} \tag{6.105}$$

For $\Omega(x)$ the simplest $S_3 \to S_3$ map is

$$\Omega^{(0)}(x) = 1 \tag{6.106}$$

More interesting is

$$\Omega^{(1)}(x) = \frac{x_4 + i\mathbf{x}\cdot\boldsymbol{\sigma}}{r} \tag{6.107}$$

belonging to the more general class

$$\Omega^{(\nu)}(x) = \big(\Omega^{(1)}(x)\big)^\nu \tag{6.108}$$

Substitution of Eq. (6.105) into Eq. (6.101) using Eqs. (6.94) and (6.95) gives

$$\nu = \frac{1}{24\pi^2}\epsilon_{\mu\alpha\beta\gamma}\int_{S_1} d^3S_\mu\, \mathrm{Tr}\big(\partial_\alpha\Omega\Omega^{-1}\partial_\beta\Omega\Omega^{-1}\partial_\gamma\Omega\Omega^{-1}\big) \tag{6.109}$$

integrated over $S_3$.

Let us evaluate this for $\Omega^{(1)}(x)$ of Eq. (6.107). Now

$$\Omega^{(1)-1} = \frac{x_4 - i\mathbf{x}\cdot\boldsymbol{\sigma}}{r} \tag{6.110}$$

We may take a unit hypersphere and since by hyperspherical symmetry the integrand in Eq. (6.109) is constant, we evaluate it at one point: $x_4 = 1$, $x = 0$, where the index $\mu = 4$. Hence at this point

$$(\partial_i\Omega)\Omega^{-1} = -i\sigma_i \tag{6.111}$$

Also,

$$\epsilon^{ijk}\,\mathrm{Tr}\big(\partial_i\Omega\Omega^{-1}\partial_j\Omega\Omega^{-1}\partial_k\Omega\Omega^{-1}\big) = 6i\,\mathrm{Tr}(\sigma_1\sigma_2\sigma_3) = 12 \tag{6.112}$$

The hyperspheric area is $2\pi^2$, so that Eq. (6.109) gives $\nu = 1$ for the winding number. For combinations $\Omega = \Omega_1\Omega_2$ the corresponding $\nu = \nu_1 + \nu_2$ for the transformations of, for example, Eq. (6.108).

Let us be more explicit for the simplest $\nu = 1$ instanton case. As $r \to \infty$, we need

$$A_\mu \to -\frac{i}{g}\big(\partial_\mu\Omega^{(1)}\big)\Omega^{(1)-1} \tag{6.113}$$

and straightforward algebra gives

$$A_\mu^a \to \frac{2}{g}\frac{\eta_{a\mu\nu}x_\nu}{r^2} \tag{6.114}$$

where we use the symbol [78]

$$\eta_{a\mu\nu} = \begin{cases} \epsilon_{aij} & 1 \leqslant \mu,\ \nu \leqslant 3 \\ -\delta_{a\nu} & \mu = 4 \\ +\delta_{a\mu} & \nu = 4 \end{cases} \tag{6.115}$$

Now Eq. (6.114) is singular at $r = 0$ and is hence unacceptable as it stands. So we try the hyperspherically symmetric ansatz

$$A_\mu^a = \frac{2}{g}\frac{\eta_{a\mu\nu}x_\nu}{r^2}f(r) \tag{6.116}$$

where $f(\infty) = 1$ and $f(0) = 0$.

For guidance in finding $f(r)$, consider the fact that

$$\int d^4x\,\mathrm{Tr}(F_{\mu\nu} - \mathscr{F}_{\mu\nu})^2 \geqslant 0 \tag{6.117}$$

which means that the energy satisfies

$$E = \int d^4x \, \mathrm{Tr}(F_{\mu\nu} F_{\mu\nu}) \geqslant 8\pi^2 |\pi| \tag{6.118}$$

This lower bound is saturated if the configuration is self-dual

$$F_{\mu\nu} = \mathscr{F}_{\mu\nu} \tag{6.119}$$

and for this case, the classical equations of motion are satisfied since the energy is stationary. Substitution of the ansatz equation (6.116) in the self-duality condition, Eq. (6.119), gives

$$rf' = 2f(1 - f) \tag{6.120}$$

with solution

$$f(r) = \frac{r^2}{r^2 + \rho^2} \tag{6.121}$$

giving the one-instanton formula

$$A^a_\mu = \frac{2}{g} \frac{\eta_{a\mu\nu} x_\nu}{r^2 + \rho^2} \tag{6.122}$$

This really has five parameters: four for the position chosen here as $x_\mu = 0$ and one for the size parameter $\rho$. For multi-instanton configurations, the number of parameters becomes $(8\nu - 3)$ since each instanton has a position (4), size (1), and gauge orientation (3), but the overall gauge choice (3), unlike the relative choices, has no physical significance. Such more general configurations have been analyzed extensively (e.g., Refs. [79–83]).

Having solved the U(1) problem by instantons, we now discuss the strong CP problem created by instantons. We have a complicated vacuum structure with candidate vacua $|\nu\rangle$, in all of which $F_{\mu\nu} = 0$. The gauge transformation $\Omega^{(1)}(x)$ has a corresponding unitary operator $T$, which takes

$$T|\nu\rangle = |\nu + 1\rangle \tag{6.123}$$

Now $T$ commutes with the Hamiltonian $H$ since it is a gauge transformation. Since $T$ is unitary, its eigenvalues are $e^{-i\theta}$ $(0 \leqslant \theta < 2\pi)$. The true vacuum eigenstates are hence the $\theta$-vacua [77, 84, 85]

$$|\theta\rangle = \sum_{\nu=-\infty}^{\infty} e^{i\nu\theta} |\nu\rangle \tag{6.124}$$

$$T|\theta\rangle = e^{-i\theta} |\theta\rangle \tag{6.125}$$

As we shall see, the parameter $\theta$ can have physical significance despite its absence from the defining QCD Lagrangian.

In the $\theta$-basis, the vacuum persistence amplitude is diagonal:

$$\langle\theta'|e^{-iHt}|\theta\rangle = \delta(\theta-\theta')\sum_{n,n'} e^{i(n-n')}\langle n'|e^{iHt}|n\rangle \tag{6.126}$$

We may write a path integral formulation in Euclidean space as

$$\langle\theta'|e^{-iHt}|\theta\rangle = \delta(\theta-\theta')I(\theta) \tag{6.127}$$

where

$$I(\theta) = \sum_{\nu} e^{-i\nu\theta}\int DA_\mu \exp\left(-\int d^4x L\right) \tag{6.128}$$

$$= \int DA_\mu \exp\left[-\int d^4x(L+\Delta L)\right] \tag{6.129}$$

$$\Delta L = \frac{i\theta}{16\pi^2}\,\mathrm{Tr}(F_{\mu\nu}\mathscr{F}_{\mu\nu}) \tag{6.130}$$

This promotes $\theta$ to a parameter in the Lagrangian; the fact that $\Delta L$ is a total derivative does not render it irrelevant because of the nontrivial instanton contribution to surface terms.

Now, $\Delta L$ violates time reversal T and parity P invariances (the action of charge conjugation C is trivial since gluons are self-conjugate). In particular, it breaks the symmetry of CP. If the quarks (or at least one quark) had been massless, we could have made a chiral rotation on the massless quarks(s),

$$q_k \to e^{i\alpha\gamma_5}q_k \tag{6.131}$$

leading to a change in the action of

$$\delta S = -\alpha\int d^4x\left(\partial_\mu J^5_\mu\right) \tag{6.132}$$

which means that

$$\delta S = 2n_f\alpha \tag{6.133}$$

Thus we may arrange $\alpha$ such that $\delta\theta = -\theta$, which means that CP is conserved by strong interactions.

But if all quarks are massive, we must consider

$$\mathscr{L} = -\frac{1}{4}\mathrm{Tr}(F_{\mu\nu}F_{\mu\nu}) + \bar{q}_k \not{D} q_k - \bar{q}_k M_{kl} q_l - \frac{i\theta}{16\pi^2}\mathrm{Tr}(F_{\mu\nu}\mathscr{F}_{\mu\nu}) \tag{6.134}$$

Now any chiral rotation such as Eq. (6.131) will alter the mass matrix phase, but in the diagonal basis,

$$-\bar{q}_k|m_k|q_k \tag{6.135}$$

the mass matrix must be real. We may then define $\bar{\theta}$ by

$$\bar{\theta} = \theta - \arg\det M \tag{6.136}$$

and are left with some $\bar{\theta}$, which leads to strong CP violation. The most sensitive limit is from the neutron electric dipole moment $d_n \lesssim 10^{-26}$ electron-cm which requires that [86, 87]

$$\theta < 10^{-10} \tag{6.137}$$

This is the strong CP problem: Why is $\bar{\theta}$ is QCD so small?

One logical possibility is that $m_u = 0$, but we have seen (Chapter 4) that one seems to require $m_u \simeq 4.5$ MeV phenomenologically, zero being quite out of accord with the analysis of pseudoscalar Goldstone boson masses.

What appears to be necessary is to arrange, as first suggested by Peccei and Quinn [88, 89], that the combined QCD and electroweak theory possess a color-anomalous global $U(1)_{PQ}$ axial symmetry, even for all quarks massive. This can allow $\bar{\theta} = 0$, restoring strong CP conservation. The simplest such model is to take *two* Higgs doublets, which give masses separately to up and down quarks. In the notation of Chapter 4, we write the Yukawa terms

$$\sum_{m,n}\left(\Gamma^u_{mn}\bar{q}_{mL}\phi_1 u_{nR} + \Gamma^d_{mn}\bar{q}_{mL}\phi_2 d_{nR}\right) + \text{h.c.} + V(\phi_1, \phi_2) \tag{6.138}$$

and hence impose, including on $V(\phi_1, \phi_2)$, the $U(1)_{PQ}$ symmetry

$$u_R \to e^{-2i\alpha}u_R \tag{6.139a}$$

$$d_R \to e^{i\alpha}d_R \tag{6.139b}$$

$$\phi_1 \to e^{2i\alpha}\phi_1 \tag{6.140}$$

$$\phi_2 e^{-i\alpha}\phi_2 \tag{6.141}$$

This color-anomalous symmetry allows one to rotate $\bar{\theta}$ back to zero at an absolute minimum of $V$ [88, 89]. The $\phi_1$ and $\phi_2$ develop vacuum values

$$\langle\phi_1\rangle = \left\langle\begin{pmatrix}\phi_1^0\\ \phi_1^+\end{pmatrix}\right\rangle = \begin{pmatrix}0\\ V_u/\sqrt{2}\end{pmatrix} \tag{6.142}$$

$$\langle\phi_2\rangle = \left\langle\begin{pmatrix}\phi_2^+\\ \phi_1^0\end{pmatrix}\right\rangle = \begin{pmatrix}0\\ V_d/\sqrt{2}\end{pmatrix} \tag{6.143}$$

such that $V_u^2 + V_d^2 = V^2$, with $V = 2^{-1/4}G^{-1/2} \simeq 248$ GeV, giving the normal $W^\pm$ and $Z^0$ masses.

There is a problem, however, with this simple model: the spontaneous breaking of $\mathrm{U(1)_{PQ}}$ implies the existence of a pseudo-Goldstone boson—the axion—with a mass of a few 100 keV [90, 91] with prescribed couplings to fermions and to photons; such an axion has been excluded experimentally.

Fortunately, there are several ways to alter the axion properties [92–96] while solving the strong CP problem. In Ref. [95], for example, the Peccei–Quinn model described above is extended to include one additional color and electroweak singlet scalar $\phi$, which does not couple to fermions but is introduced into $V(\phi_1, \phi_2, \phi)$ such that there is symmetry as in Eqs. (139) through (6.141), together with

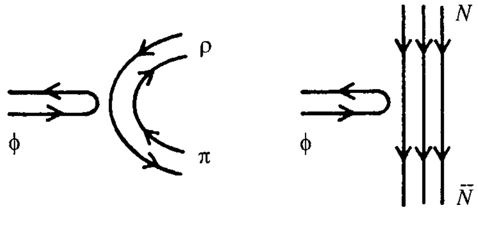
$$\phi \to e^{-2i\alpha}\phi \tag{6.144}$$

If it is arranged that $\langle\phi\rangle = V_H \gg V$, the axion mass can be reduced by the ratio $(V/V_H)$ and, even more important, its coupling to ordinary matter can simultaneously be reduced by $(V/V_H)$. This "invisible axion" is then hard to exclude experimentally, although its detection has been discussed [97, 98]. Present 2008 data suggest that, if the axion exists, its mass lies between one microvolt (μeV) and one millivolt (meV).

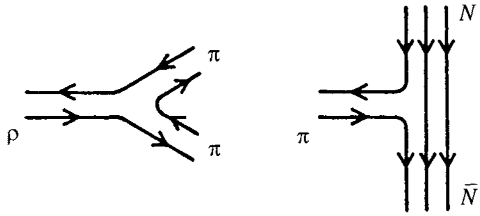
## 6.5
## $1/N$ Expansion

Because QCD is lacking in expansion parameters—the perturbative expansion in the coupling constant is very limited in usefulness, as we have seen—it is desirable to introduce such a parameter from outside, 't Hooft [99, 100] made a fruitful suggestion of the use of the parameter $1/N$, where $N$ is the number of colors; one should take the limit $N \to \infty$ and expand in $1/N$ around that limit with the assumption that $N = 3$ is not too different from $N \to \infty$.

Here we shall discuss how the $1/N$ expansion enables us to understand qualitatively certain striking phenomenological features of strong interactions: in particular, the absence of exotic mesons with structure $(qq\bar{q}\bar{q})$ compared to the abundance of $(q\bar{q})$ meson bound states. Also, we can understand successfully the validity of the Okubo–Zweig–Iizuka (OZI) rule [101–103], which suppresses couplings such as $\phi\rho\pi$ and $\phi NN$ (Fig. 6.10*a* and *b*) relative to, say, $\rho\pi\pi$ and $\pi NN$ (Fig. 6.11*a* and *b*) because the former diagrams, unlike the latter, involve the annihilation of a quark–antiquark pair with the same meson (the $\phi$ meson), resulting in a quark diagram that is divisible into two pieces unconnected by quark lines. For decay of heavy charmonium or bottomium mesons, the OZI rule can be attributed to asymptotic freedom giving small coupling, but for the lighter mesons, the $1/N$ expansion provides the simplest explanation.




(a) (b) **Figure 6.10** Suppressed couplings.




(a) (b) **Figure 6.11** Allowed couplings.

The QCD theory with $N$ colors involves $N$ Dirac fields $\psi^a$ $(1 \leqslant a \leqslant N)$ and $(N^2 - 1)$ gluon fields $A_\mu{}^a{}_b$ with $A_\mu{}^a{}_b = 0$. The field strength is

$$F_{\mu\nu}{}^a{}_b = \partial_\mu A_\nu{}^a{}_b - \partial_\nu A_\mu{}^a{}_b + i\left(A_\mu{}^a{}_c A_\nu{}^c{}_b - A_\nu{}^a{}_c A_\mu{}^c{}_b\right) \tag{6.145}$$

and the Lagrangian is written in the form

$$L = \frac{N}{g^2}\left[-\frac{1}{4}F_{\mu\nu}{}^a{}_b F_{\mu\nu}{}^b{}_a + \bar{\psi}_a\left(i\partial_\mu A_\mu{}^a{}_b\right)\gamma_\mu \psi^b - m\bar{\psi}_a\psi^a\right] \tag{6.146}$$

For convenience the $N/g^2$ has been scaled out. Putting $A = A'(\sqrt{N}/g)$ and $\psi = \psi'(\sqrt{N}/g)$ would reestablish the coupling parameter as $g/\sqrt{N}$, which turns out to be the unique arrangement that makes the $1/N$ expansion useful. We shall let $N \to \infty$ with $g^2 N$ fixed, with a fixed number $(N_f)$ of flavors. A slightly different limit $N \to \infty$, $N_f \to \infty$ with $g^2 N$ and $g^2 N_f$ fixed has been advocated by Veneziano [104], but here we shall consider only the 't Hooft limit.

It is useful to keep account of color indices in Feynman graphs by setting up the double-line [105–107] representation; whereas a quark line has a single index propagating according to

$$\left\langle \psi^a(x)\bar{\psi}_b(y)\right\rangle = \delta^a_b S(x - y) \tag{6.147}$$

The gluon propagator has two indices propagating as in

$$\left\langle A_\mu{}^a{}_b(x) A_\nu{}^c{}_d(y)\right\rangle = \left(\delta^a_d \delta^c_b - \frac{1}{N}\delta^a_b \delta^c_d\right) D_{\mu\nu}(x - y) \tag{6.148}$$

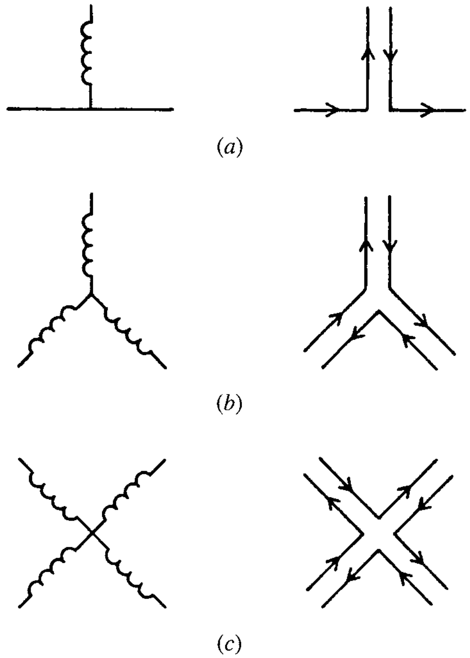


Figure 6.12 Double-line formalism.

In leading order of the $1/N$ expansion, we may drop the trace term in Eq. (6.148) and then, as far as the indices go, the gluon propagator is like a quark–antiquark pair and hence may be represented by a double line. Some examples of the double-line representation are given alongside the corresponding Feynman graph in Fig. 6.12.

Consider first vacuum-persistence Feynman graphs with no external lines. Regard each index loop as the border of a polygon; such polygons are fitted together to cover a two-dimensional surface by identifying a double-line gluon propagator with the edge of two polygons. The surface is oriented for $SU(N)$ because of the quark arrows, so we obtain only spheres with holes and handles, not Klein bottles.

Suppose that this surface associated with a vacuum-persistence graph has $V$ vertices, $E$ edges, and $F$ faces. From the Lagrangian, Eq. (6.146), we see that each vertex carries an $N$ and each edge contains a $1/N$. Each face carries an $N$ because of the sum over colors in the loop. The power of $N$ is thus given by the Euler characteristic ($\chi$):

$$N^{F+V-E} = N^{\chi} \tag{6.149}$$

If we regard the oriented surface as a sphere with $H$ handles and $B$ holes (boundaries), then $\chi$ is given by

$$\chi = 2 - 2H - B \tag{6.150}$$

This is the standard topological characteristic. To derive these formulas, consider first a sphere which can be obtained by gluing together two $n$-sided polygons at their perimeters, giving $E = V = n$ and two faces $F = 2$. This has $\chi = 2$. Now cut a hole that changes $\Delta F = -1$ and $\Delta B = +1$. To attach a handle, cut two holes in both $n$-sided polygons and identify their perimeters; this makes $\Delta F = \Delta V = -n$ and $\Delta F = -2$. This then confirms both Eqs. (6.149) and (6.150).

The leading vacuum-persistence graphs therefore are going like $N^2$ and have the "index surface" corresponding to a sphere. Any closed quark loop will generate a hole in the surface and decrease $\chi$; hence the leading graphs contain only gluons, and are proportional to $N^2$. The leading graphs that involve quarks have one closed quark loop and go like $N$, since $B = 1$.

What is also important is that these leading graphs are all planar; that is, the Feynman graph can be drawn on a plane surface with no lines crossing. This is clear because for the $N^2$ gluon graphs we may remove an arbitrary face and project the remainder of the sphere on a plane to produce a graph with a gluon around the outside corresponding to the perimeter of the removed face. For the leading quark graphs, proportional to $N$, one face is already absent, so we immediately write the planar graph encircled by the single quark loop.

Let $B_i$ be local gauge-invariant operators that have one $\psi$ and one $\bar{\psi}$ (i.e., are bilinear in the quark fields) and consider a connected Green's function for the vacuum expectation of a string of $m$ such bilinears. To maintain our $1/N$ analysis, we add sources to the action by

$$S' = S + N \sum J_i B_i \tag{6.151}$$

keeping an $N$ in each vertex, including the bilinear insertions, and then the required Green's function is obtained from $Z(j_i)$, the generating functional for connected graphs by

$$\langle B_1 \cdots B_m \rangle = \frac{1}{N^m} \frac{\delta^m}{\delta_{j_i} \cdots \delta_{j_m}} z \bigg|_{J_i=0} \tag{6.152}$$

The leading diagrams are planar and are surrounded by a single quark loop on which the bilinear insertions sit. From Eq. (6.152) each insertion gives $1/N$, so we have

$$\langle B_1 \cdots B_m \rangle \sim N^{1-m} \tag{6.153}$$

We may introduce gauge-invariant local operator $G_i$ made solely from gluons and deal with them similarly to find that

$$\langle G_1 \cdots G_n \rangle \sim N^{2-n} \tag{6.154}$$

where the insertions are now anywhere in the planar graph. For a mixture of the two types of operator, the leading term has one quark loop and goes as

$$\langle B_1 B_2 \cdots B_m G_1 G_2 \cdots G_n \rangle \sim N^{1-m-n} \tag{6.155}$$

To proceed to phenomenology we assume that QCD confines color for arbitrarily large $N$, not only for $N = 3$.

A meson is created by acting with a $B$ on the vacuum. In fact, because the couplings will be set by $1/\sqrt{N}$, we must renormalize $B_i' = \sqrt{N'} B_i$ whereupon

$$\langle B_1' \cdots B_m' \rangle \sim N^{1-(1/2)m} \tag{6.156}$$

The leading order hence behaves as if $B_i'$ were a fundamental field with fundamental coupling $1/\sqrt{N}$ and as if it were a Born approximation! One can now even argue that the only singularities at leading order in $1/N$ are poles, also like a Born approximation. To show this, consider

$$\langle B_1' B_2' \rangle \tag{6.157}$$

and prove that $B_2'$ creates only single-meson states. For if $B_2'$ created, say, two mesons which are subsequently rescattered by two further bilinears and then annihilated by a fourth bilinear—all of order $O(1)$—we would have a physical singularity at $O(1)$ in the connected Green's function for four bilinears, in contradiction to Eq. (6.153). For consistency, therefore, $B_i'$ creates only single-meson states and the only singularities are poles like a Born amplitude.

Another point for $N \to \infty$ is that asymptotic freedom is maintained (actually, it becomes stronger for fixed $N_f$). Thus the two-point function must behave logarithmically at spacelike high energy, and this is impossible unless there are an infinite number of meson poles; any finite number of poles generates power behavior.

For all of these reasons—the treelike behavior in coupling ($1/\sqrt{N}$) and the meromorphy with an infinite number of meson poles—it is often speculated that a string theory coincides with QCD in the extreme $N \to \infty$ limit, and that the appropriate duality we see for strong interactions is because $N = 3$ approximates $N \to \infty$.

The OZI rule can by now be easily understood, at least for mesons. An OZI forbidden process involves at least two closed quark loops and is hence suppressed relative to an OZI-allowed process that has contributions from one closed quark loop. For baryons, the $1/N$ expansion is more complicated [108, 109] and better approached from chiral pion models, as we shall see.

Absence of exotic mesons (e.g., $qq\bar{q}\bar{q}$) can be seen by noticing that the only way to make a quadrilinear in quark fields that is gauge invariant (color singlet) is to take the product of two bilinears which are each color singlet; this is a simple property of $SU(N)$ for large $N$. Thus we may put

$$Q(x) = B_1'(x) B_2'(x) \tag{6.158}$$

But now observe that in

$$\begin{aligned}\langle Q^+(x) Q(y) \rangle = &\langle B_1'^+(x) B_1'(y) \rangle \langle B_2'^+(x) B_2'(y) \rangle \\ &+ \langle B_1'^+(x) B_2'(y) \rangle \langle B_2'^+(x) B_1'(y) \rangle \\ &+ \langle B_2'^+(x) B_1'^+(y) B_1'(y) B_2'(y) \rangle \end{aligned} \tag{6.159}$$

the third term is suppressed by $1/N$ compared to the first two terms, which physically mean independent propagation of two nonexotic mesons. Thus, exotic mesons are suppressed by $1/N$.

All of these attractive observations can be exhibited explicitly in QCD with $N \to \infty$ for two space–time dimensions [99, 100] a theory that can be solved almost completely.

To complete the picture, we discuss the effective chiral meson Lagrangian for large $N$ QCD and how baryons are solitons (Skyrmions) in this weakly coupled meson theory.

The point is that while at zero temperature the $SU(N_f)_L \times SU(N_f)_R$ chiral symmetry of QCD for $N_f$ flavors is realized in the Nambu–Goldstone mode with massless pseudoscalar bosons in an octet for $N_f = 3$, we expect at high temperature the realization in the Wigner–Weyl mode, implying a "chiral" phase transition from broken to unbroken phase of this symmetry. This may be studied by looking at QCD from the viewpoint of the Landau theory, precisely analogous to the Landau–Ginzburg effective theory of superconductivity, which is an effective theory generated by the BCS microscopic theory.

Of course, calculation of the parameters in the effective theory from the underlying QCD theory is highly nontrivial.

The natural choice of order parameter is the $3 \times 3$ complex matrix

$$M_{ij}(x) = \sum_{\alpha=1}^{n} \bar{\psi}_i^a \frac{1}{2}(1+\gamma_5)\psi_j^\alpha(x) \tag{6.160}$$

The simplest Lagrangian for $M(x)$ incorporating the relevant symmetries is

$$\begin{aligned}\mathscr{L} &= a\,\mathrm{Tr}\big(\partial_\mu M \partial_\mu M^+\big) + b\,\mathrm{Tr}\big(M^+ M\big) + c\,\mathrm{Tr}\big(MM^+MM^+\big) \\ &\quad + d\,\mathrm{Tr}\big(MM^+\big)^2 + e\ln\det M\end{aligned} \tag{6.161}$$

The last term violates the axial U(1) symmetry as required by the anomaly

$$\partial_\mu J_\mu^{L,R} = \pm\frac{g^2}{N}\frac{1}{64\pi^2}\,\mathrm{Tr}(F_{\mu\nu}\mathscr{F}_{\mu\nu}) \tag{6.162}$$

and hence the coefficient $e$ is $O(1/N)$. It is the $\eta'$ mass term, and hence the $\eta'$ mass is of order $1/N$ relative to the pseudoscalar octet.

To make this clearer we may write $M = He^{i\phi}U$, where $U$ is unitary and $H$ is the $M$ vacuum expectation—determined by the coefficients $b$, $c$, and $d$ in Eq. (6.161)—in the Nambu–Goldstone phase. Now $\phi$ is the $\eta'$ field ($\phi = \ln\det M$) and the $\pi^a$ pion octet is contained in

$$U(x) = \exp\left[\frac{2i\pi^a(x)\lambda^a}{F_\pi}\right] \tag{6.163}$$

The massless bosons have Lagrangian

$$\mathscr{L} = \frac{1}{16} F_\pi^2 \partial_\mu U \partial_\mu U^+ \quad (6.164)$$

$$= \mathrm{Tr}(\partial_\mu \pi \partial_\mu \pi) + \frac{1}{F_\pi^2} \mathrm{Tr}\big(\partial_\mu \pi^2 \partial_\mu \pi^2\big) + \cdots \quad (6.165)$$

where we have exploited the crucial fact that $F_\pi^2 \sim N$. To see this, consider the correlation function

$$G(x) = \big\langle j_\mu^f(x) j_\mu^5(0) \big\rangle \quad (6.166)$$

where $j_\mu^5 = \bar{\psi}\gamma_\mu\gamma_5\psi$ is the axial current. At large distances only the lightest meson intermediate state survives, giving

$$G(x) \sim \big|\langle \pi | j_\mu^5(0) | 0\rangle\big|^2 e^{-m_\pi x} \quad (6.167)$$

$$\sim N \quad (6.168)$$

because the leading contribution has only one quark loop, as we have seen. Thus

$$F_\pi^2 \sim \big|\langle \pi | j_\mu^5(0) | 0\rangle\big|^2 \sim N \quad (6.169)$$

as required, and $m_\pi$ in Eq. (6.167) is $N$ independent. The chiral model of Eq. (6.165) has been used successfully in strong interactions (e.g., Ref. [110]).

In the large $N$ limit, baryon mass goes like $N$ [108], as can be expected from the one-gluon exchange energy behaving as $\frac{1}{2}N(N-1)\cdot(g^2/N)$, since the coupling is $g/\sqrt{N}$. This is suggestive of a nonperturbative soliton in the chiral theory whose coupling is $\sim (1/N)$. However, keeping only the Lagrangian in Eq. (6.161) allows no static soliton solutions, and one needs higher derivative terms. In 1961, Skyrme [111] considered the model

$$L = \frac{F_\pi^2}{16\pi^2} \mathrm{Tr}\big(\partial_\mu U \partial_\mu U^+\big) - \frac{1}{32e^2} \mathrm{Tr}\big[\partial_\mu U U^+, \partial_\nu U U^+\big]^2 \quad (6.170)$$

and showed that this contains topological entities (Skyrmions) which might be identified with baryons. The value of $U(x)$ tends to a constant for large $x$ in the broken symmetry phase where $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R \to \mathrm{SU}(3)_{L+R}$, so we are, at fixed time, mapping $S_3$ into SU(3) using $\pi_3(\mathrm{SU}(3)) = Z$, allowing a definition of baryon number by

$$B = \frac{i}{24\pi^2} \int d\mathbf{x}\, \mathrm{Tr}\big(\epsilon_{ijk} \partial_i U U^+ \partial_j U U^+ \partial_k U U^+\big) \quad (6.171)$$

The Skyrmion solution with $B = 1$ has the form

$$U(\mathbf{x}) = \exp\big[i\hat{\mathbf{x}} \cdot \boldsymbol{\sigma} f(r)\big] \quad (6.172)$$

with $f(r)$ behaving as in Fig. 6.13, such that $f(0) = \pi$, $f(\infty) = 0$.

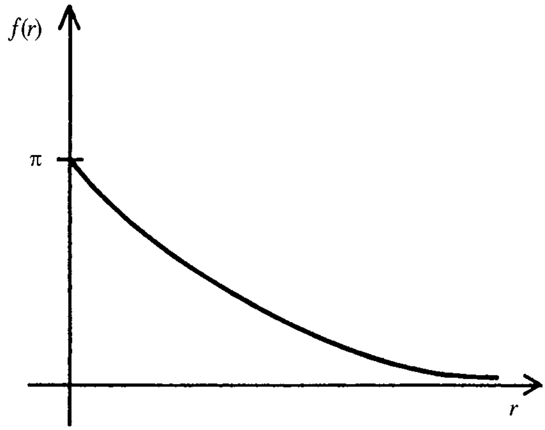



**Figure 6.13** Behavior of $f(r)$.

Further, more recent examination [112–116] has shown that the Skyrmions are fermions only for $N$ = odd (e.g., $N = 3$) [113] and has even given some qualitative agreements [114] with the baryon spectrum observed.

## 6.6 Lattice Gauge Theories

Because of the peculiar feature of the color force that it is weak at short distance and strong at large distance, we need a quite different approach to examine QCD at low energy. Such an approach is fortunately provided by the lattice method of regularization suggested in 1974 by Wilson [117]. One treats QCD on a discretized Euclidean four-space–time, similarly to the statistical treatment of many-body systems. Confinement is automatic on the lattice, and numerical studies in 1979 by Creutz [118–120] made very plausible that no phase transition interrupts the connection from strong-coupled confinement to weak-coupled asymptotic freedom. The numerical studies allow one to relate the QCD scale ($\Lambda$) to the string tension between quarks, to the mass gap (or glueball mass), and even to hadron masses. The limit is not now experimental accuracy (as it is for perturbative QCD and $e^+e^-$ annihilation), but the size and speed of electronic computers. Our treatment will be a brief overview—detailed reviews are available (e.g., Refs. [121] and [122]).

The continuum formulation we wish to regularize is that in Euclidean space–time (time $t \to it$) with action $S(\phi)$ depending on fields $\phi$ and certain masses $m$ and couplings $g$. The quantum expectation value of observable $O(\phi)$ is

$$\langle O(\phi)\rangle = Z^{-1}\int D\phi\, O(\phi) e^{-S(\phi)} \tag{6.173}$$

where

$$Z = \int D\phi e^{-S(\phi)} \tag{6.174}$$

is the vacuum-persistence amplitude, analogous to the partition function in the statistical formulation of a thermodynamic system. At the end we may wish to continue $\langle O(\phi)\rangle$ back to Minkowski space–time.

Consider a hypercubic space–time lattice with sites labeled by $n$ and basis vectors $\mu$ ($1 \leqslant \mu \leqslant 4$) of length $a$, the lattice spacing. Before discussing gauge fields, we shall warm up with matter fields. The action for free scalars

$$S = \int d^4x \left[\frac{1}{2}(\partial_\mu \phi)^2 + \frac{1}{2}\left(m^2\phi^2\right)\right] \tag{6.175}$$

becomes

$$S = a^4 \sum_{n,\mu} \left(\frac{\phi_{n+\mu} - \phi_n^2}{2a^2} + \frac{m^2}{2}\phi_n^2\right) \tag{6.176}$$

The lattice sites $n = a(n_1, n_2, n_3, n_4)$, where $n_i$ are integers satisfying $-N/2 < n_i \leqslant +N/2$ on an $N^4$ lattice with the identification of $n_i$ with $(n_i + N)$ for periodic boundary conditions. $N$ provides an infrared cutoff just as the lattice spacing ($a$) provides an ultraviolet cutoff.

The integral in Eq. (6.173) now becomes an ordinary multiple integral according to

$$D\phi \to \prod_n d\phi_n \tag{6.177}$$

The partition function becomes

$$Z = \int \prod_n d\phi_n e^{-S} = \left(\det \frac{M}{2\pi}\right)^{-1/2} \tag{6.178}$$

where $M$ is defined by

$$S = \phi_m M_{mn} \phi_n \tag{6.179}$$

To gain familiarity, and to savor the resemblance to solid-state theory, we take the lattice Fourier transform of any $f_n$ to be

$$\tilde{f}_k = \sum_n f_n e^{2\pi i k_\mu n_\mu / N} \tag{6.180}$$

with inverse

$$f_n = n^{-4} \sum_k \tilde{f}_k e^{-2\pi i k_\mu n_\mu / N} \tag{6.181}$$

Using relations such as

$$\sum_n f^*_{n+\mu} g_n = N^{-4} \sum_k f^*_k q_k e^{2\pi i k_\mu / N} \tag{6.182}$$

we may rewrite the kinetic energy term in Eq. (6.176) as

$$a^4 \sum_{n,\mu} (\phi^*_{n+\mu}\phi_{n+\mu} + \phi^*_n \phi_n - \phi^*_{n+\mu}\phi_n - \phi^*_n \phi_{n+\mu})$$

$$= \frac{a^2}{2N^4} \sum_{k,\mu} \left(2 - 2\cos \frac{2\pi k_\mu}{N}\right) |\tilde{\phi}_k|^2 \tag{6.183}$$

and hence the full action is

$$S = a^4 N^{-4} \sum_k \frac{1}{2} \mathscr{M}_k |\tilde{\phi}|^2 \tag{6.184}$$

$$\mathscr{M}_k = m^2 + 2a^{-2} \sum_\mu \left(1 - \cos \frac{2\pi k_\mu}{N}\right) \tag{6.185}$$

Consider now the scalar propagator. It is given by

$$\langle \phi_m \phi_n = \left(M^{-1}\right)_{mn} = \frac{1}{a^4 N^4} \sum_k \mathscr{M}_k^{-1} e^{2\pi i k\cdot(m-n)/N} \tag{6.186}$$

Taking the limit $N \to \infty$ and setting

$$q_\mu = \frac{2\pi k_\mu}{aN} \tag{6.187}$$

$$\frac{1}{a^4 N^4} \sum_k \to \int \frac{d^4 q}{(2\pi)^4} \tag{6.188}$$

$$x_\mu = -a_n(m_\mu - n_\mu) \tag{6.189}$$

gives

$$\langle \phi_m \phi_n \rangle = \int_{-\pi/a}^{+\pi/a} \frac{d^4 q}{(2\pi)^4} \frac{e^{-iq\cdot x}}{M^2 + 2a^{-2}\sum_\mu (1 - \cos q a_\mu)} \tag{6.190}$$

If we take the lattice spacing to zero, then by expanding the cosine we regain the continuum propagator from Eq. (6.190).

If we follow the parallel steps for a spin-$\frac{1}{2}$ field, we encounter what is called the "fermion-doubling" problem: In fact, on a four-dimensional hypercubic lattice, a direct treatment leads to 16 replicas of the continuum fermion. Let us see how this happens and how it has been handled.

The action for free fermions

$$S = \bar{\psi}(\not\partial + m)\psi \tag{6.191}$$

leads to

$$\mathcal{M}_k = m + \frac{i}{a}\sum_{\mu}\gamma_\mu \sin\frac{2\pi k_\mu}{N} \tag{6.192}$$

and to the propagator

$$\langle\psi_m\bar{\psi}_n\rangle = \int_{-\pi/a}^{+\pi/a}\frac{d^4q}{(2\pi)^4}\frac{e^{-iq\cdot x}}{m + (i/a)\sum_\mu \gamma_\mu \sin aq_\mu} \tag{6.193}$$

Equation (6.193) for fermions has a problem not shared by Eq. (6.190) for scalars. If we take the limit $a \to 0$ in Eq. (6.193), there are contributions from the integral end point. To proceed we put $q\!\!\!/_\mu = q_\mu - \pi/a$ and rewrite (for each dimension)

$$\int_{-\pi/a}^{+\pi/a} dq_\mu = \int_{-\pi/2a}^{+\pi/2a} dq_\mu + \int_{-\pi/a}^{+\pi/a} dq\!\!\!/_\mu \tag{6.194}$$

Each piece of the integration region gives a free fermion propagator in the continuum limit, so we have $2^4 = 16$ "flavors" of fermion. Clearly, this is intimately related to the fact that the field equation is first order rather than second order. It may also be understood in terms of the triangle anomaly—a single chiral fermion gives an anomaly in the continuum theory which must be canceled in the lattice version, where the ultraviolet divergence is regulated; the lattice achieves this by generating pairs of fermions with opposite chirality [123]. Only one pair would be necessary for this, however, and the large number 16 is due to the choice of a hypercubic lattice; a minimum of two flavors has also been deduced by Nielsen and Ninomiya considering a general lattice [124].

To avoid this species doubling, the most common procedure is due to Wilson [125], who modified the action to give

$$\mathcal{M}_k = m + \frac{i}{a}\sum_{\mu}\gamma_\mu \sin aq_\mu + \frac{r}{a}\sum_{\mu}(1 - \cos aq_\mu) \tag{6.195}$$

Here $r$ is a free parameter. For $q_\mu$ near zero the extra term is of the order of the lattice spacing and does not affect the continuum limit. For $q_\mu$ near $\pi/a$ the new term leads to

$$i\not q + \left(m + \frac{2r}{a}\right) \tag{6.196}$$

and so acts like a mass term $(2r/a)$. This breaks chiral symmetry even when $m = 0$ and hence raises a problem in studying chiral properties of QCD. One may also try to remove the large $q_\mu$ components by distributing different spinor components on different lattice sites, making effectively smaller sublattices [126, 127].

Now consider a pure Yang–Mills theory with gauge group G. When we recall that quantities such as

$$\phi^+(x+dx)e^{ig\int a_\mu^\alpha \tau_\alpha \, dx_\mu}\phi(x) \tag{6.197}$$

are locally gauge invariant in the continuum theory, it is natural to think of gauge transformations as transport operators and to associate independent matrices $U_{ij}$ belonging to G with each *link* of the lattice between nearest-neighbor sites.

The link is oriented and we clearly need

$$U_{ji} = U_{ij}^{-1} \tag{6.198}$$

Under a gauge transformation, the matter fields transform as

$$\phi_i \to G_i \phi_i \tag{6.199}$$

while for the transport operators,

$$U_{ij} \to G_i U_{ij} G_j^{-1} \tag{6.200}$$

so that $\phi_j^+ U_{ji} \phi_i$ is gauge invariant.

Now consider a path $(\gamma)$ through the lattice $i_1 \to i_2 \to \cdot \to i_N$. The corresponding transport operator

$$U_j = U_{i_N i_{N-1}} U_{i_{N-1} i_{N-2}} \cdots U_{i_2 i_1} \tag{6.201}$$

transforms as

$$U_j \to G_{i_N} U_j G_{i_1}^{-1} \tag{6.202}$$

In particular, consider a closed path $(\lambda)$ with operator $U_\lambda$, which transforms as

$$U_\lambda \to G_{i_1} U_\lambda G_{i_1}^{-1} \tag{6.203}$$

This means that the trace of $U_\lambda$ is gauge invariant:

$$W_\lambda = \mathrm{Tr}(U_\lambda) \tag{6.204}$$

This is called a Wilson loop [117].

The field strength tensor $F_{\mu\nu}^\alpha$ of the continuum theory corresponds to a curl and hence a line integral around a small closed contour. Thus consider the path around an elementary square or *plaquette* $(P)$ of the hypercubic lattice

$$U_P = U_{i_1 i_4} U_{i_4 i_3} U_{i_3 i_2} U_{i_2 i_1} \tag{6.205}$$

For SU$(N)$ we define the lattice action for a plaquette as

$$S_P = \frac{1}{g^2}\left[1 - \frac{1}{2N}\,\mathrm{Tr}(U_P)\right] \tag{6.206}$$

and the action for the full finite lattice is then

$$S = \sum_P S_P \tag{6.207}$$

The continuum limit of the classical action is easily seen to be

$$S = \int d^4x\left(-\frac{1}{4}F^\alpha_{\mu\nu}F^\alpha_{\mu\nu}\right) \tag{6.208}$$

For example, with SU(2), we may write

$$U_{ji} = \exp\left(igA^\alpha_\mu\frac{\sigma^\alpha}{2}a\right) \tag{6.209}$$

$$= 1\cos\theta_{ji} + i\boldsymbol{\sigma}\cdot\mathbf{n}_{ij}\sin\theta_{ji} \tag{6.210}$$

where

$$\theta_{ji} = \frac{1}{2}ga\left|A^\alpha_\mu\right| \tag{6.211}$$

Similarly for the plaquette

$$U_P = \exp\left(igaF^\alpha_{\mu\nu}\frac{\sigma^\alpha}{2}\right) \tag{6.212}$$

$$= 1\cos\theta_P + i\boldsymbol{\sigma}\cdot\mathbf{n}_P\sin\theta_P \tag{6.213}$$

where

$$\theta_P = \frac{1}{2}ga^2\left|F^\alpha_\nu\right| \tag{6.214}$$

For the plaquette action we take

$$\frac{1}{g^2}\left(1 - \frac{1}{2}\cos\theta_P\right) = \frac{1}{g^2}\left[1 - \frac{1}{4}\,\mathrm{Tr}(U_P)\right] \tag{6.215}$$

and then in the sum of Eq. (6.207) replace

$$a^4\sum \to \int d^4x \tag{6.216}$$

to obtain Eq. (6.208) as the continuum limit.

There are three remarks concerning the lattice $S$:

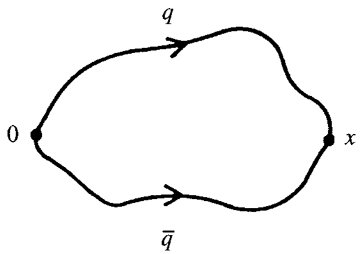



**Figure 6.14** Wilson loop.

1. The action regularized by the lattice is gauge invariant, and because it involves a finite sum, no gauge fixing is necessary. On the lattice we can forget about, for example, Faddeev–Popov ghosts.
2. Because the basic dynamical variables are finite group elements ($U_{ij}$), we may equally consider discrete gauge groups. Indeed, the study of, for example, $Z_N$ gauge groups has played a significant role in the history of this subject [128, 129].
3. The choice of *hypercubic* lattice is only for convenience but is almost universally used in applications. The idea of a lattice with random sites [130–132] has been proposed with a view to retaining perhaps more of the rotational and Lorentz symmetry properties which are exact in the continuum.

Our next topic is the strong-coupling expansions of the lattice gauge theory. Let us consider $\mathrm{SU}(N)$ with matter fields (quarks) in the defining representation, and consider the vacuum expectation of a Wilson loop, $W_\lambda$. We take the quarks as external sources $\psi(x)$ and consider the quantity

$$\left\langle \bar{\psi}(x)\Gamma\psi(x)\bar{\psi}(0)\Gamma\psi(0)\right\rangle_\lambda \tag{6.217}$$

corresponding to creation of a quark–antiquark pair later annihilating at $x$ (Fig. 6.14). When we quantize we make the weighted average of the Wilson loop operator $W_\lambda$:

$$W_\lambda = \exp\left[ig\int_\gamma A_\mu(x)\,dS_\mu\right] \tag{6.218}$$

which on the lattice is a sum of the form

$$\exp\left[\sum_\gamma(\pm)\theta_{ij}\right] \tag{6.219}$$

This sign is $(\pm)$ according to the direction of each link. Inserting in the discretized version of Eq. (6.173) with the action $S$ given by Eq. (6.207), we have

$$\langle W_\lambda\rangle = Z^{-1}\prod_{i,j}\int_{-\pi}^{+\pi} d\theta_{ij}\exp\left[i\sum_\gamma(\pm\theta_{ij}) + \frac{1}{g^2}\sum_P\left(1-\frac{1}{2}\cos\theta_P\right)\right] \tag{6.220}$$

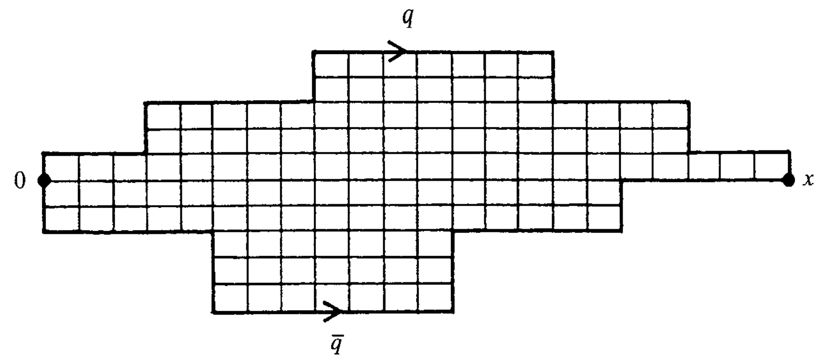



**Figure 6.15** Tiling of Wilson loop.

$$Z = \prod_{i,j} \int_{-\pi}^{+\pi} d\theta_{ij} \exp\left[\frac{1}{g^2} \sum_P \left(1 - \frac{1}{2}\cos\theta_P\right)\right] \tag{6.221}$$

Now expand $[Z\langle W_\lambda \rangle]$ in powers of $(1/g^2)$

$$Z\langle W_\lambda \rangle = \sum_{k=0}^{\infty} W_\lambda^{(k)} \left(G^2\right)^{-k} \tag{6.222}$$

The zeroth-order term is

$$W_\lambda^{(0)} = \prod_{i,j} \int_{-\pi}^{+\pi} d\theta_{ij} \exp\left[i \sum_\gamma (\pm\theta_{ij})\right] \tag{6.223}$$

which vanishes since

$$\int_{-\pi}^{+\pi} dx \exp(\pm ix) = 0 \tag{6.224}$$

The next term ($k = 1$) will also vanish by periodicity unless $\lambda$ is precisely a single plaquette. It is clear that the first nonvanishing term will be where we may "tile" the Wilson loop with elementary plaquettes (Fig. 6.15). This occurs when $k = A/a^2$, $A$ being the minimal area enclosed by the path $\lambda$; all the interior links cancel in this term because they each occur twice, once in each direction. This we find that

$$\langle W_\lambda \rangle \sim \exp\left(-\frac{A \ln g^2}{a^2}\right) \tag{6.225}$$

This area law is the signal for confinement due to a linear potential. Consider a rectangular loop $\lambda$; then for a linear potential, we expect an action $\sim KRT$, where $K$ is the string tension, parallel to that of string models (e.g., Ref. [133]).

Higher terms in $1/g^2$ correspond to different evolutions of the "string" for a given path $\gamma$ where the area of the "world sheet" is nonminimal. For large $g^2$, therefore, we have a confining theory, and we may picture the field lines between quark and antiquark bunched together into a confining string rather than a Maxwellian inverse-square distribution.

For the continuum limit of Eq. (6.225), we need to keep

$$\frac{\ln g^2}{a^2} = \text{constant} \tag{6.226}$$

to avoid a singularity, but this requirement is inconsistent with the $g^2 \gg 1$ assumption already made. Thus we cannot take the continuum limit after making the strong-coupling expansion.

This is actually very fortunate because the confining picture we have reached in the strong-coupling expansion on the lattice holds for the Abelian case like QED and we know that QED does not confine electrons! Thus, the whole issue of whether QCD confines hinges on whether we can connect smoothly between the strong-coupling regime and the scaling regime of weak coupling. There must be no intervening phase transition. There must be one is QED. These were the speculations of Wilson [114], who provided, however, no supporting evidence in 1974. Such evidence came later from Monte Carlo simulations, which attempt to connect the strong-coupling regime to the scaling regime.

Let us discuss the continuum limit of the quantum theory. As $a \to 0$, the coupling must be renormalized in a very specific way. First and foremost, the limit must correspond to a scaling critical point as follows. Let $q_i$ by physical observables with dimension $(-d_i)$ in length. Then

$$q_1 = a^{-d_1} f_1(g) \tag{6.227}$$

$$q_2 = a^{-d_2} f_2(g) \tag{6.228}$$

and so on. For correlation lengths $l_1$,

$$l_1 = a f_1(g) \tag{6.229}$$

$$a_2 = a f_2(g) \tag{6.230}$$

and so on. Putting $a \to 0$ will not give a meaningful limit unless *all* $f_i(g) \to \infty$ $(d_i > 0)$ or $\to 0$ $(d_i < 0)$. It is not obvious that this is possible; it requires $g$ to approach a critical point $g = g_{\text{cr}}$ as $a \to 0$.

Introducing a mass scale $\Lambda_L$ associated with the lattice regularization, we may set

$$a = \frac{1}{\Lambda} f(g) \tag{6.231}$$

and $f(g)$ must set, by universal scaling, all the $f_i(g)$ according to

$$f_i(g) \simeq c_i \big(f(g)\big)^{d_i} \tag{6.232}$$

This can be seen from the fact that

$$a(g) \simeq q_i^{-1/d_i} f_i(g)^{1/d} \tag{6.233}$$

for all $i$, as $g \to g_{\text{cr}}$. Also,

$$q_i = c_i \Lambda^{d_i} \tag{6.234}$$

The function $f(g)$ follows from our knowledge that

$$-a\frac{dg}{da} = +\beta_1 g^3 + \beta_2 g^5 + O\left(g^7\right) \tag{6.235}$$

where $\beta_1$ and $\beta_2$ are the one- and two-loop $\beta$-function coefficients derived in Chapter 5. For $\mathrm{SU}(3)$,

$$\beta_1 = \frac{11}{16\pi^2} \tag{6.236}$$

$$\beta_2 = \frac{102}{(16\pi^2)^2} \tag{6.237}$$

To two loops the solution of Eq. (6.235) is easily seen to be

$$(\Lambda_L a) = f(g) = \exp\left(-\frac{1}{2\beta_1 g^2}\right)\left(g^2\beta_1\right)^{\beta_2/2\beta_1} \tag{6.238}$$

The lattice scale $\Lambda_L$ can be related to the scale $\Lambda$ evaluated phenomenologically in deeply inelastic scattering [134, 135]. Note that the correct relation between $g$ and $a$ is not at all as suggested by Eq. (6.226), where we interchanged strong-coupling and continuum limits.

There remains the problem of connecting up the strong-coupling and scaling regions. This cannot be done analytically. Also, it cannot be done exactly even for a $10^4$ lattice, for example, because even with gauge group $Z_2$, the number of links is 40,000 (number of links $= d \times$ number of sites in $d$ dimensions), and the number of terms in the partition function is

$$2^{40,000} \simeq 1.58 \times 10^{12041} \tag{6.239}$$

It will *never* be possible to add this number of terms exactly! Thus we must approximate by a stochastic procedure which selects a large sample of configurations with probability distribution proportional to $\exp(-S)$; the exact quantum average in Eq. (6.173) is approximated by averages taken over the large sample of configurations. This is called a Monte Carlo simulation because it involves selection of a random number in a fashion similar to that of a roulette wheel.

In a Monte Carlo simulation, one begins with some configuration $C^{(1)}$ and generates by some procedure configurations $C^{(2)}, \ldots, C^{(k)}$ such that as $k$ grows, the probability approaches $\exp[-S(C)]$. If after $n_0$ steps we are sufficiently near the limit, a quantum average may be approximated by

$$\langle 0 \rangle = \frac{1}{n} \sum_{k=n_0+1}^{n_0+n} O\left(C^{(k)}\right) \tag{6.240}$$

Let $p(C \to C')$ be the transition probability matrix. Then the Boltzmann distribution $\exp[-S(C)]$ must be an eigenvector of this matrix. A sufficient (but not necessary) condition for this is *detailed* balance

$$\frac{p(C \to C')}{p(C' \to C)} = \frac{e^{-S(C')}}{e^{S(C)}} \tag{6.241}$$

If we use the stochastic property

$$\sum_{C'} p(C - C') = 1 \tag{6.242}$$

it follows from Eq. (6.241) that

$$\sum_{C} e^{-S(C)} p(C \to C') = e^{S(C')} \tag{6.243}$$

as required.

The detailed procedure most frequently used is that of Metropolis et al. [136]. One starts with some symmetric distribution

$$p_0(C \to C') = p_0(C' \to C) \tag{6.244}$$

and considers a change to $C'$ from $C$ by it, with the corresponding change in action

$$\Delta(S) = S(C') - S(C) \tag{6.245}$$

A random number $(r)$ is chosen (roulette wheel) between 0 and 1. Then of $r < e^{\Delta S}$, the change in configuration is accepted; otherwise, it is rejected. That is, the new configuration is accepted with conditional probability $e^{-\Delta S}$. If $\Delta S$ is negative, the change is always accepted.

This algorithm satisfies the detailed balance condition since

$$\frac{p(C \to C')}{p(C' \to C)} = \frac{p_0(C \to C')}{p_0(C' \to C)} e^{-\Delta S} = \frac{e^{-S(C')}}{e^{-S(C)}} \tag{6.246}$$

Using the Metropolis procedure, one systematically upgrades each link $U_{ij}$ through the lattice (one "iteration"). It often saves time to upgrade each link several times since the rejection rate is high; this is the *improved Metropolis algorithm*. An infinite number of upgrades per link is the *heat bath method*, but this is usually too slow. A large number of iterations is made until the relevant quantum average is stabilized. The simplest and perhaps most fundamental calculation is of the string tension. Considering $(R \times T)$ rectangular Wilson loops, one may calculate [120]

$$x = \ln \frac{W(R, T)W(R-1, T-1)}{W(R-1, T)W(R, T-1)} \tag{6.247}$$

which cancels all perimeters and constant pieces in the action, leaving only the string tension $x$ in the strong-coupled domain. More generally,

$$W(R, T) \to \exp\left[-E_0(R)T\right] \tag{6.248}$$

where $T \to \infty$, where $E_0(R)$ is the minimum energy state, and hence

$$s = a^2 \frac{dE_0(R)}{dR} \tag{6.249}$$

gives the force between quarks for all $R$.

In this way, one can examine whether there is a phase transition between strong and weak coupling. By successively heating up, starting from temperature zero (all $U_{ij} = 1$), then cooling down from infinite temperature (all $U_{ij}$ = random), the appearance of hysteresis signals a phase transition. In 1979, Creutz [118] showed convincingly that whereas SU(2) in $d = 5$ dimensions and SO(2) in $d = 4$ dimensions did show hysteresis, SU(2) in $d = 4$ did not. Instead, the $d = 4$ SU(2) case showed no structure except a rapid crossover from strong to weak coupling. These results are shown in Fig. 6.16. The lattice size used there was an amazingly small $5^4$ for $d = 4$ and $4^5$ for $d = 5$.

A next step [119, 120] was to relate the string tension at strong coupling known from Regge slopes to be about (400 MeV)$^2$ [or (14 tons)$^2$] to the scale $\Lambda_L$ and hence [134, 135] to $\Lambda_{\overline{\text{MS}}}$. The result was $\Lambda_{\overline{\text{MS}}} \simeq 200 \pm 35$ MeV, in agreement with the value derived from perturbative calculations of scaling in semihadronic processes.

Many efforts have been made (e.g., Refs. [137–141]) to establish the mass gap of pure QCD (i.e., the lightest glueball mass). This involves studying the correlation function

$$G(x) = \langle O(x)O(0)\rangle - \langle O(x)\rangle\langle O(0)\rangle \tag{6.250}$$

$$x \xrightarrow{\sim} \infty e^{-m|x|} \tag{6.251}$$

for appropriate local operators $O(x)$, where $m$ is the lightest glueball mass. Results suggest that $m$ is around 1.5 GeV.

Finally, and most significantly, since the experimental data are the best quality, there are the hadron masses [141–146]. Again one studies correlation functions such as Eq. (6.250) where the local operator is, for example, $O(x) = \bar{\psi}(x)\gamma_\mu\psi(x)$ for a vector meson. In such computations, virtual quark loops are difficult to handle and often ignored. Nevertheless, the results for meson and baryon masses (with quark masses as input) agree with experiment to within a few percent.

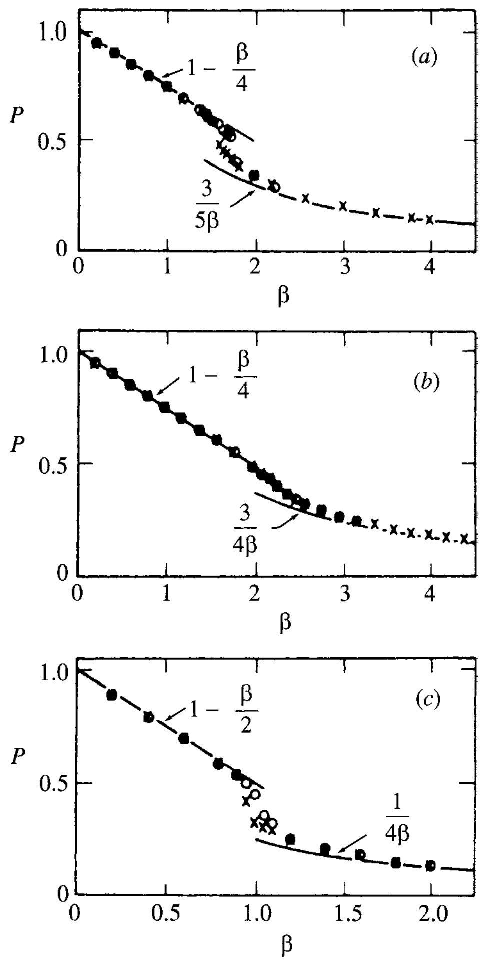



**Figure 6.16** (a) SU(2) in $d = 5$; (b) SU(2) in $d = 4$; (c) SO(2) in $d = 4$. [After M. Creutz, Phys. Rev. Lett. 43, 554 (1979).]

## 6.7 Summary

There remains no doubt that QCD is the correct fundamental theory of strong interactions up to energies of 100 GeV. The frustration is in finding a single prediction that is reliable to better than 0,1% accuracy. To some extent we have been spoiled by the spectacular success of QED, where $(g - 2)$ agrees to 1 part in $10^{10}$.

Nature has nevertheless been generous to the theorist since the asymptotically free regime of perturbative QCD sets in at very low scales $Q^2 \gtrsim (2\ \text{GeV})^2$. Also, lattice QCD with very small lattices having only $10^4$ sites (or even less) works better than one might reasonably expect.

It may be some time before extremly high precision comparison of QCD to experiment becomes commonplace. One needs more accurate and higher-energy experimental data and much larger and faster electronic computers to improve on estimations of hadron masses and other static properties.

The area of "hadronic physics" is now a large research field. Although not germane to discovering new fundamental laws of physics, it is comparable to nuclear physics. The role of protons and neutrons in the nuclei are replaced in it by the next level: quarks and gluons as constituents of mesons and baryons.

## References


**1** H. Fritzsch and M. Gell-Mann, in *Proceedings of the 16th International Conference on High Energy Physics*, Chicago, 1972, Vol. 2, p. 135.

**2** S. Weinberg, *Phys. Rev. Lett.* 31, 494 (1973).

**3** D.J. Gross and F. Wilczek, *Phys. Rev.* D8, 3633 (1973).

**4** H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Phys. Lett.* 47B, 365 (1973).

**5** Y. Nambu, in *Preludes in Theoretical Physics*, A. De Shalit, ed., North-Holland, New York, 1966.

**6** M. Gell-Mann, *Phys. Lett.* 8, 214 (1964).

**7** G. Zweig, CERN report 8419/TH 412, unpublished, 1964.

**8** D.B. Lichtenberg and S.P. Rosen, eds., *Developments in the Quark Model of Hadrons*, Vol. I, 1964–1978, Hadronic Press, Nonantum, Mass., 1980, p. 22.

**9** O.W. Greenberg, *Phys. Rev. Lett.* 13, 598 (1964).

**10** M.Y. Han and Y. Nambu, *Phys. Rev.* 139, B1005 (1965).

**11** D.J. Gross and F. Wilczek, *Phys. Rev. Lett.* 26, 1343 (1973).

**12** H.D. Politzer, *Phys. Rev. Lett.* 26, 1346 (1973).

**13** W.E. Lamb and R.C. Retherford, *Phys. Rev.* 72, 241 (1947).

**14** P. Kusch and H. Foley, *Phys. Rev.* 72, 1256 (1947).

**15** P. Kusch and H. Foley, *Phys. Rev.* 73, 412 (1948).

**16** T. Appelquist and H. Georgi, *Phys. Rev.* D8, 4000 (1973).

**17** A. Zee, *Phys. Rev.* D8, 4038 (1973).

**18** G. 't Hooft, *Nucl. Phys.* B61, 455 (1973).

**19** W.A. Bardeen, A.J. Buras, D.W. Duke, and T. Muta, *Phys. Rev.* D18, 3998 (1978).

**20** E.G. Floratos, D.A. Ross, and C.T. Sachrajda, *Nucl. Phys.* B129, 66 (1977) [erratum: *Nucl. Phys.* B139, 545 (1978)].

**21** R. Jost and J. Luttinger, *Helv. Phys. Acta* 23, 201 (1950).

**22** M. Dine and J. Sapirstein, *Phys. Rev. Lett.* 43, 668 (1979).

**23** K.G. Chetyrkin and A.L. Kataev, *Phys. Lett.* 85B, 277 (1979).

**24** W. Celmaster and R.J. Gonzales, *Phys. Rev. Lett.* 42, 1435 (1979).

**25** W. Celmaster and R.J. Gonzales, *Phys. Rev.* D20, 1420 (1979).

**26** W. Celmaster and R.J. Gonzales, *Phys. Rev. Lett.* 44, 560 (1980).

**27** W. Celmaster and R.J. Gonzales, *Phys. Rev.* D21, 3112 (1980).

**28** C.H. Llewellyn Smith, in *High Energy Physics, 1980*, L. Durand and L.G. Pondrum, eds., AIP Publishing, New York, 1981, p. 134.

**29** P.M. Stevenson, *Phys. Rev.* D23, 2916 (1981).

**30** P.M. Stevenson, *Phys. Rev.* D24, 1622 (1981).

**31** M.R. Pennington, *Phys. Rev.* D26, 2048 (1982).

**32** P.M. Stevenson, *Nucl. Phys.* B203, 472 (1982).

**33** J.C. Wrigley, *Phys. Rev.* D27, 1965 (1983).

**34** P.M. Stevenson, *Nucl. Phys.* B231, 65 (1984).

**35** A. Dhar, *Phys. Lett.* 128B, 407 (1983).

**36** C.J. Maxwell, *Phys. Rev.* D28, 2037 (1983).

**37** C.J. Maxwell, *Phys. Rev.* D29, 2884 (1984).

**38** F.J. Dyson, *Phys. Rev.* 85, 631 (1952).

**39** E.R. Whittaker and G.N. Watson, *Modern Analysis*, 4th ed., Cambridge University Press, Cambridge, 1927.

**40** L.N. Lipatov, *JETP Lett.* 25, 104 (1977).

**41** G. 't Hooft, in *The Whys of Subnuclear Physics*, Erice School 1977, A. Zichichi, ed., Plenum Publishing, New York, 1979, p. 943.
**42** G. Parisi, *Phys. Lett.* 66B, 167 (1977).
**43** G. Parisi, *Phys. Lett.* 66B, 382 (1977).
**44** E. Brezin, J.C. Le Guillou, and J. Zinn-Justin, *Phys. Rev.* D15, 1544, 1558 (1977).
**45** W.Y. Crutchfield, *Phys. Rev.* D19, 2370 (1979).
**46** S.D. Drell, D.J. Levy, and T.-M. Yan, *Phys. Rev.* 187, 2159 (1969).
**47** S.D. Drell, D.J. Levy, and T.-M. Yan, *Phys. Rev.* D1, 1617 (1970).
**48** N. Cabibbo, G. Parisi, and M. Testa, *Nuovo Cimento Lett.* 4, 35 (1970).
**49** R.P. Feynman, *Phys. Rev. Lett.* 23, 1415 (1969).
**50** J.D. Bjorken and E.A. Paschos, *Phys. Rev.* 185, 1975 (1969).
**51** G. Hanson et al., *Phys. Rev. Lett.* 35, 1609 (1975).
**52** Ch. Berger et al., *Phys. Lett.* 78B, 176 (1978).
**53** Ch. Berger et al., *Phys. Lett.* 81B, 410 (1979).
**54** F. Bloch and H. Nordsieck, *Phys. Rev.* 52, 54 (1937).
**55** T. Kinoshita and A. Sirlin, *Phys. Rev.* 113, 1652 (1959).
**56** T. Kinoshita, *J. Math. Phys.* 3, 650 (1962).
**57** T.D. Lee and M. Nauenberg, *Phys. Rev.* 133, B1549 (1964).
**58** J. Ellis, M.K. Gaillard, and G.G. Ross, *Nucl. Phys.* B111, 253 (1976).
**59** T.A. DeGrand, Y.J. Ng, and S.H.H. Tye, *Phys. Rev.* D16, 3251 (1977).
**60** G. Sterman and S. Weinberg, *Phys. Rev. Lett.* 39, 1436 (1977).
**61** A. De Rujula, J. Ellis, E.G. Floratos, and M.K. Gaillard, *Nucl. Phys.* B138, 387 (1978).
**62** P.M. Stevenson, *Phys. Lett.* 78B, 451 (1978).
**63** B.G. Weeks, *Phys. Lett.* 81B, 377 (1979).
**64** P. Hoyer, P. Osland, H.G. Sander, T.F.'Walsh, and P.M. Zerwas, *Nucl. Phys.* B161, 349 (1979).
**65** G. Grunberg, Y.J. Ng, and S.H.H. Tye, *Nucl. Phys.* B168, 315 (1980).
**66** A. Ali, *Phys. Lett.* 110B, 67 (1982).
**67** R. Brandelik et al., *Phys. Lett.* 86B, 243 (1979).
**68** D.P. Barber et al., *Phys. Rev. Lett.* 43, 830 (1979).
**69** Ch. Berger et al., *Phys. Lett.* 86B, 418 (1979).
**70** D.P. Barber et al., *Phys. Lett.* 89B, 139 (1979).
**71** T.D. Lee, *Particle Physics and Introduction to Field Theory*, Harwood Academic, London, 1981, pp. 746–768.
**72** J.D. Bjorken and S. Brodsky, *Phys. Rev.* D1, 1416 (1970).
**73** H. Georgi and M. Machacek, *Phys. Rev. Lett.* 39, 1237 (1977).
**74** E. Farhi, *Phys. Rev. Lett.* 39, 1587 (1977).
**75** A.A. Belavin, A.M. Polyakov, A.S. Schwartz, and Yu.S. Tyupkin, *Phys. Lett.* 59B, 85 (1975).
**76** S. Weinberg, *Phys. Rev.* D11, 3583 (1975).
**77** G. 't Hooft, *Phys. Rev. Lett.* 37, 8 (1976).
**78** G. 't Hooft, *Phys. Rev.* D14, 3432 (1976).
**79** E. Witten, *Phys. Rev. Lett.* 38, 121 (1977).
**80** R. Jackiw, C. Nohl, and C. Rebbi, *Phys. Rev.* D15, 1642 (1977).
**81** M.F. Atiyah and R.S. Ward, *Commun. Math. Phys.* 55, 117 (1977).
**82** M.F. Atiyah, N.J. Hitchin, V.G. Drinfeld, and Yu.I. Manin, *Phys. Lett.* 65A, 185 (1978).
**83** E.F. Corrigan, D.B. Fairlie, R.G. Yates, and P. Goddard, *Commun. Math. Phys.* 58, 223 (1978).
**84** R. Jackiw and C. Rebbi, *Phys. Rev. Lett.* 37, 172 (1976).
**85** C.G. Callan, R.F. Dashen, and D.J. Gross, *Phys. Lett.* 63B, 334 (1976).
**86** V. Baluni, *Phys. Rev.* D19, 227 (1979).
**87** R. Crewther, P. Di Vecchia, G. Veneziano, and E. Witten, *Phys. Lett.* 88B, 123 (1979).
**88** R.D. Peccei and H.R. Quinn, *Phys. Rev. Lett.* 38, 1440 (1977).
**89** R.D. Peccei and H.R. Quinn, *Phys. Rev.* D16, 1791 (1977).
**90** S. Weinberg, *Phys. Rev. Lett.* 40, 223 (1978).
**91** F. Wilczek, *Phys. Rev. Lett.* 40, 279 (1978).
**92** D.M. McKay and H. Muczek, *Phys. Rev.* D19, 985 (1978).
**93** J.E. Kim, *Phys. Rev. Lett.* 43, 103 (1979).
**94** M. Shifman, A. Vainshtein, and V. Zakharov, *Nucl. Phys.* B166, 493 (1980).
**95** M. Dine, W. Fishcler, and M. Srednicki, *Phys. Lett.* 104B, 199 (1981).

96 H.P. Nilles and S. Raby, *Nucl. Phys.* B198, 102 (1982).
97 P. Sikivie, *Phys. Rev. Lett.* 51, 1415 (1983).
98 P. Sikivie, *Phys. Rev. Lett.* 52, 695 (1983).
99 G. 't Hooft, *Nucl. Phys.* B72, 461 (1974).
100 G. 't Hooft, *Nucl. Phys.* B75, 461 (1974).
101 S. Okubo, *Phys. Lett.* 5, 165 (1963).
102 G. Zweig, CERN preprint, unpublished, 1964.
103 J. Iizuka, *Suppl. Prog. Theor. Phys.* 37/38, 21 (1966).
104 G. Veneziano, *Nucl. Phys.* B117, 519 (1976).
105 H. Harari, *Phys. Rev. Lett.* 22, 562 (1969).
106 J. Rosner, *Phys. Rev. Lett.* 22, 689 (1969).
107 S. Coleman, Lectures at Erice Summer School, 1979, in *Pointlike Structure Inside and Outside Hadrons*, A. Zichichi, ed., Plenum Publishing, New York, 1982, p. 11.
108 E. Witten, *Nucl. Phys.* B160, 57 (1979).
109 E. Corrigan and P. Ramond, *Phys. Lett.* 87B, 73 (1979).
110 B.W. Lee, *Chiral Dynamics*, Gordon and Breach, New York, 1971.
111 T.H.R. Skyrme, *Proc. R. Soc.* A260, 127 (1961).
112 A.P. Balachandran, V.P. Nair, S.G. Rajeev, and A. Stern, *Phys. Rev. Lett.* 49, 1124 (1982).
113 E. Witten, *Nucl. Phys.* B233, 422, 433 (1983).
114 G.W. Adkins, C.R. Nappi, and E. Witten, *Nucl. Phys.* B228, 552 (1983).
115 A. Dhar and S. Wadia, *Phys. Rev. Lett.* 52, 959 (1984).
116 A. Dhar, R. Shankar, and S. Wadia, *Phys. Rev.* D31, 3256 (1985).
117 K.G. Wilson, *Phys. Rev.* D10, 2445 (1974).
118 M. Creutz, *Phys. Rev. Lett.* 43, 553 (1979).
119 M. Creutz, *Phys. Rev.* D21, 2308 (1980).
120 M. Creutz, *Phys. Rev. Lett.* 45, 313 (1980).
121 M. Creutz, *Quarks, Gluons and Lattices*, Cambridge University Press, Cambridge, 1983.
122 C. Rebbi, *Lattice Gauge Theories and Monte Carlo Simulations*, World Scientific, Singapore, 1983.
123 L.H. Karsten and J. Smit, *Nucl. Phys.* B183, 103 (1981).
124 H.B. Nielsen and M. Ninomiya, *Nucl. Phys.* B185, 20 (1981).
125 K.G. Wilson, in *New Phenomena in Subnuclear Physics*, A. Zichichi, ed., Plenum Publishing, New York, 1977.
126 J. Kogut and L. Susskind, *Phys. Rev.* D11, 395 (1975).
127 T. Banks, S. Raby, L. Susskind, J. Kogut, D.R.T. Jones, P.N. Scharbach, and D.K. Sinclair, *Phys. Rev.* D15, 1111 (1977).
128 M. Creutz, L. Jacobs, and C. Rebbi, *Phys. Rev. Lett.* 42, 1390 (1979).
129 M. Creutz, L. Jacobs, and C. Rebbi, *Phys. Rev.* D20, 1915 (1979).
130 N.H. Christ, R. Friedberg, and T.D. Lee, *Nucl. Phys.* B202, 89 (1982).
131 N.H. Christ, R. Friedberg, and T.D. Lee, *Nucl. Phys.* B210, [FS6], 310 (1982).
132 N.H. Christ, R. Friedberg, and T.D. Lee, *Nucl. Phys.* B210, [FS6], 337 (1982).
133 P.H. Frampton, *Dual Resonance Models*, W.A. Benjamin, New York, 1974.
134 A. Hasenfratz and P. Hasenfratz, *Phys. Lett.* 93B, 165 (1980).
135 R. Dashen and D.J. Gross, *Phys. Rev.* D23, 2340 (1981).
136 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.* 21, 1087 (1953).
137 B. Berg, A. Billoire, and C. Rebbi, *Ann. Phys.* 142, 185 (1982).
138 M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi, F. Rapuano, B. Taglienti, and Y.-C. Zhang, *Phys. Lett.* 110B, 295 (1982).
139 B. Berg and A. Billoire, *Phys. Lett.* 113B, 65 (1982).
140 K. H. Mutter and K. Schilling, *Phys. Lett.* 117B, 75 (1982).
141 H. Hamber and G. Parisi, *Phys. Rev. Lett.* 47, 1792 (1981).
142 E. Marinari, G. Parisi, and C. Rebbi, *Phys. Rev. Lett.* 47, 1795 (1981).
143 H. Hamber, E. Marinari, G. Parisi, and C. Rebbi, *Phys. Lett.* 108B, 314 (1982).
144 D. Weingarten, *Phys. Lett.* 109B, 57 (1982).
145 A. Hasenfratz, Z. Kunszt, P. Hasenfratz, and C.B. Lang, *Phys. Lett.* 110B, 289 (1982).
146 F. Fucito, G. Martinelli, C. Omero, G. Parisi, R. Petronzio, and F. Rapuano, *Nucl. Phys.* B210, 407 (1982).

# 7
# Model Building

## 7.1
## Introduction

In the first edition, it seemed logical at that time (1986) to include a final chapter on the subject of grand unification. That idea was first proposed in the 1970s and became very popular in the first half of the 1980s. Since the principal prediction of grand unification, proton decay, remains unconfirmed, I have decided to step back in the final chapter of the second edition and look first at what puzzles are presented by the standard model, then give examples of motivated and testable models that address some of the issues. For this third edition, I have extended this further to include an additional chapter, Chapter 8, on model building.

In Section 7.2 the questions posed by the theory are listed; this provides a worksheet for the model builder aiming to go beyond the established theory. We then present four illustrative examples which may inspire the reader to try to make further examples and perhaps find the correct high-energy theory. It should be obvious that more data are needed at present (2008) to discriminate between possibilities and the LHC will likely provide these. In Section 7.3 the left–right model is discussed based on the notion that parity may be restored as a symmetry at high energy. In Section 7.4 we describe chiral color theory, where the strong interaction QCD undergoes a spontaneous breaking analogous to that of the electroweak theory. In Section 7.5 a possible explanation for the occurrence of three families is offered by the 331-model. In Section 7.6 we use inspiration from the duality between string theory and field theory to study the possibility that the strong and electroweak interactions become conformal at TeV energies. This last topic is discussed more fully in Chapter 8.

## 7.2
## Puzzles of the Standard Model

The standard strong/electroweak theory (SM) has successfully fit all reproducible data. The model was built in the 1960s by Glashow [1] Salam [2], and Weinberg [3].

It was shown to be renormalizable by 't Hooft in 1971 [4]. Its experimental verification was already in excellent shape by the time of the 1978 Tokyo Conference [5]. The $W^{\pm}$ and $Z^0$ were discovered in 1983. From 1983 until the present there have been various ambulances to chase, where data disagreed with the SM, but further data analysis has so far always rescued the SM.

Nevertheless, the SM has its own shortcomings and incompleteness, which motivate model-building "beyond the SM." A good starting point is to examine critically the large number of parameters that must be fit phenomenologically in the SM.

The gauge sector of the SM is based on the group

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \tag{7.1}$$

and includes 12 gauge bosons with 24 helicity states.

The fermions occur in the three families taking massless neutrinos

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \bar{u}_L \bar{d}_L \quad \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \bar{e}_L \tag{7.2}$$

$$\begin{pmatrix} c \\ s \end{pmatrix}_L \bar{c}_L \bar{s}_L \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \bar{\mu}_L \tag{7.3}$$

$$\begin{pmatrix} t \\ b \end{pmatrix}_L \bar{t}_L \bar{b}_L \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \bar{\tau}_L \tag{7.4}$$

It requires 45 fields to describe these fermions. The scalars are in the complex doublet

$$\begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} \tag{7.5}$$

with four fields. The total number of fields in the SM is therefore 73. Perturbatively, the baryon and lepton numbers $B$ and $L$ are conserved and the neutrino masses are vanishing $M(\nu_i) = 0$.

As a model builders' worksheet, we can list the 19 parameters of the SM:

| | |
|---|---|
| 6 | Quark masses |
| 3 | Lepton masses |
| 3 | Mixing angles $\theta_i$ |
| 1 | Phase $\delta$ |
| 1 | QCD $\bar{\Theta}$ |
| 3 | Coupling constants |
| 2 | Higgs sector |
| 19 | Total |

The first 12 parameters may be addressed by horizontal symmetries, the next two are associated with models of CP violation, and the three couplings can be related, for example, in grand unification. The two Higgs parameters—the Higgs mass and the magnitude of its quartic self-coupling (which sets the size of the weak scale)—need an even bigger theoretical framework.

There are other questions unanswered in the SM framework, such as:

1. Why are there three families?
2. Why use SU(3) × SU(2) × U(1)?
3. Why are there the particular fermion representations?
4. Is the Higgs boson elementary?

Model building beyond the SM is simultaneously:

1. Trivial mathematically in the sense that the rules for building renormalizable gauge field theories have been well known since 1971.
2. Impossible physically since in the absence of experimental data departing from the SM, how can one discriminate between models?

**Supersymmetry.** The most popular model beyond the SM in unquestionably supersymmetry. Its motivation is to ameliorate (not solve) the *gauge hierarchy problem*—that, assuming a desert up to $\sim 10^{16}$ GeV, nothing stabilizes the ratio $(M_W/M_{\text{GUT}})^2 \sim 10^{-28}$. It is clearly *testable* by the prediction of a large number of particles with masses below 1 TeV.

The models described here have their motivation outside supersymmetry. But any renormalizable gauge theory can be promoted to an $N = 1$ globally supersymmetric theory. The motivation for supersymmetry arises at present from mathematical physics and not from phenomenology.

It does seem on aesthetic grounds that supersymmetry is likely to be used by nature in its fundamental theory. The question is at what scale and how supersymmetry is broken. It may, for example, be broken at the Planck scale, as in some versions of supergravity and superstrings.

## 7.3 Left–Right Model

One of the simplest and most attractive extensions of the SM, also one of the oldest [6–8], is the left–right model. It has two principal motivations: (1) to restore parity at high energies and (2) to replace the assignment of weak hypercharge by the assignment of the more familiar quantity $(B - L)$, where $B$ and $L$ are baryon number and lepton number.

The left–right model (sometimes called the Pati–Salam model) has definite testability. It predicts nonzero neutrino masses $m(\nu_i) \neq 0$. It predicts certain $\Delta B = 2$ processes such as $NN \to$ pions. Also, there are $\Delta L \neq 0$ processes, such as neutrinoless double beta decay $(\beta\beta)_{0\nu}$.

The fundamental theory is assumed to be parity symmetric. The electroweak gauge group is promoted from

$$SU(2)_L \times U(1)_Y \tag{7.6}$$

to

$$SU(2)_L \times SU(2)_R \times U(1)_{B-L} \tag{7.7}$$

The usual relation between electric charge $Q$ and $SU(2)_L$,

$$Q = T_{3L} + \frac{Y}{2} \tag{7.8}$$

becomes

$$Q = T_{3L} + T_{3R} + \frac{B-L}{2} \tag{7.9}$$

as can be seen easily by writing

$$\begin{pmatrix} u \\ d \end{pmatrix}_R \frac{1}{2} Y = \begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} T_{3R} + \frac{B-L}{2} \\ T_{3R} + \frac{B-L}{2} \end{pmatrix} \tag{7.10}$$

Similarly,

$$\begin{pmatrix} N \\ e \end{pmatrix}_R \frac{1}{2} Y = \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} T_{3R} + \frac{B-L}{2} \\ T_{3R} + \frac{B-L}{2} \end{pmatrix} \tag{7.11}$$

So we need to gauge the quantity $(B - L)$ rather than the more perplexing weak hypercharge $Y$.

The symmetry is broken in the stages

$$SU(2)_L \times SU(2)_R \times U(1)_{B-L} \times P \tag{7.12}$$

$$\xrightarrow{M_P} \quad SU(2)_L \times SU(2)_R \times U(1)_{B-L} \tag{7.13}$$

$$\xrightarrow{M_{W_R}} SU(2)_L \times U(1)_Y \tag{7.14}$$

$$\xrightarrow{M_{W_L}} U(1)_Q \tag{7.15}$$

It is often assumed that $M_P = M_{W_R}$. If there is a range where $M_P > \mu > M_{W_R}$, within it $g_{2L} \neq g_{2R}$ but $W_L$ and $W_R$ remain massless.

A minimal Higgs sector contains the scalars

$$\Delta_L(1, 0, +2) \qquad \Delta_R(0, 1, +2) \qquad \Phi\left(\tfrac{1}{2}, \tfrac{1}{2}, 0\right) \tag{7.16}$$

For a range of parameters, this gives a $P$-violating minimum.

For phenomenology, precision tests of the SM require that $M(W_R)$ and $M(Z_R)$ be greater than 500 GeV. These lower bounds come from analysis of

$$\bar{p}p \to \mu^+\mu^- + X \tag{7.17}$$

$$\mu^- \to e^-\bar{\nu}_e\nu_\mu \tag{7.18}$$

$$n \to pe^-\bar{\nu}_e \tag{7.19}$$

and so on.

$(V - A)$ theory was prompted by $m(\nu) = 0$ and $\gamma_5$ invariance. The presence of $(V + A)$ is linked to $m(\nu) \neq 0$ by seesaw formulas such as

$$m(\nu)_e \simeq \frac{m_e^2}{M(W_R)} \tag{7.20}$$

The $N_R$ state is necessary in the left–right model, just as in an SO(10) GUT.

It is natural to expect $\Delta L \neq 0$ Majorana masses and related processes such as neutrinoless double beta decay $(\beta\beta)_{0\nu}$, which has $\Delta L = 2$, $\Delta B = 0$.

Looking again at the formula

$$Q = T_{3L} + T_{3R} + \frac{B - L}{2} \tag{7.21}$$

and bearing in mind that for $E \gg M_{W_L}$ one has $\Delta I_{3L} = 0$ while always $\Delta Q = 0$, we see that the fact that $\Delta I_{3R}$ is integer implies that $|\Delta(B - L)|$ is a multiple of 2.

Thus we expect not only processes with $\Delta L = 2$ and $\Delta B = 0$ as in the $(\beta\beta)_{0\nu}$ already considered, but also processes with $\Delta L = 0$ and $\Delta B = 2$. To understand these baryon-number violating processes, it is instructive to consider partial unification to the group

$$\mathrm{SU}(4)_C \times \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R \tag{7.22}$$

[the Pati–Salam group, a subgroup of SO(10)]. Here the $\mathrm{SU}(4)_C$, where the lepton number is the "fourth" color, is broken to $\mathrm{SU}(3)_C \times \mathrm{U}(1)_{B-L}$. The $B$ violating processes are then seen to be, for example, $N\bar{N}$ oscillations and processes such as $NN \to n's$, induced by introducing a new scale of a few PeV. This multi-PeV scale is consistent with nucleon stability because the relevant operator is dimension 10 rather than dimension 6 as in SU(5) GUTs.

In summary, the key features of the left–right model are:

1. Nonvanishing $m(\nu)$ might be interpreted as evidence for the $L{-}R$ extension of the SM.
2. $\Delta B = 2$, $\Delta L = 0$ processes are expected, as are $(\beta\beta)_{0\nu}$, and so on.

## 7.4 Chiral Color

There is a marked asymmetry in the SM between the strong color interactions, which are described by an unbroken SU(3) gauge group, and the electroweak interactions, described by a gauge group $SU(2) \times U(1)$ broken spontaneously at a few hundred GeV (the weak scale) to the electromagnetic U(1).

Chiral color [9, 10] is a model where the strong interactions are more similar to the electroweak interactions. The chiral color gauge group is $SU(3)_L \times SU(3)_R$, broken at some scale (the simplest choice is the weak scale, although this is not essential) to the diagonal subgroup color SU(3).

In this case quantum chromodynamics is a relic of a spontaneous breakdown of a larger gauge group, and the breakdown leads to the existence of many new fundamental particles, including especially the *axigluon*. The axigluon is a spin-1 massive particle which should weigh several hundred GeV and be visible as a jet–jet resonance at hadron colliders. The fermion representation must be free of triangle anomalies (see Chapter 3), and this requires the existence of additional fermions. Aside from the familiar anomalies of the standard model there are potential new anomalies of the form $[SU(3)_L]^3 - [SU(3)_R]^3$, and $Q[SU(3)_L]^2 - Q[SU(3)_R]^2$. These anomalies can be avoided by a variety of tricks in which the $SU(3) \times SU(3)$ assignments of quarks are juggled and in which fermions with exotic color quantum number are introduced.

In Ref. [9] a list of anomaly-free fermion representations was presented and reproduced here. The list of five cases was intended to be illustrative but not exhaustive. The five models are very different from one another. However, there are three common threads: the necessary existence of *more* than three families of quarks and leptons, a massive octet of spin-1 hadrons (the axigluons) which are strongly coupled to quarks, and a rich spectrum of scalar mesons required for symmetry breaking.

**Mark I.** Here we simply let two of the four families transform under $SU(3)_L$, the other two under $SU(3)_R$. Unlike the subsequent models, in Mark I the anomalies are trivially canceled. Specifying the transformation behavior of left-handed chiral fermions under $[SU(3)_L, SU(3)_R, 0]$, we have four colored weak doublets:

$$2\left[\left(3, 1, \tfrac{2}{3}\right) + \left(3, 1, -\tfrac{1}{3}\right)\right] + 2\left[\left(1, 3, \tfrac{2}{3}\right) + \left(1, 3, -\tfrac{1}{3}\right)\right] \tag{7.23}$$

eight colored singlets which complete four quark families:

$$2\left(\bar{3}, 1, -\tfrac{2}{3}\right) + 2\left[\left(\bar{3}, 1, \tfrac{1}{3}\right) + 2\left[\left(1, \bar{3}, -\tfrac{2}{3}\right) + 2\left(1, \bar{3}, \tfrac{1}{3}\right)\right]\right] \tag{7.24}$$

and four charged leptons and their neutrinos.

This model is not chiral. The axigluon couples to

$$(\bar{u}\gamma_\mu u + \bar{c}\gamma_\mu c + \bar{d}\gamma_\mu d + \bar{s}\gamma_\mu s) - (\bar{t}\gamma_\mu t + \bar{h}\gamma_\mu h + \bar{b}\gamma_\mu b + \bar{l}\gamma_\mu l) \tag{7.25}$$

where the fourth family includes a h(igh) and l(ow) quark. Mass mixing leads to axigluon couplings to flavor-changing neutral currents. The angle $\phi$ connecting the lighter two families to the heavier ones must be small. GIM violation in the $ds$ and $cu$ systems is $\sim \phi^4$, while in the $bs$ and $bd$ systems it is $\sim \phi^2$.

A minimal set of scalar mesons includes $\phi(1, 1, 2)$, which can give mass to all quarks and leptons, and $\phi(3, \bar{3}, 2)$, which breaks chiral color and provides mixing between the heavier and lighter quark families.

**Mark II.** This model involves three conventional fermion families, an extra $Q = \frac{2}{3}$ quark, and an SU(3) sextet fermion or quix. There are three colored weak doublets:

$$3\left[\left(3, 1, \tfrac{2}{3}\right) + \left(3, 1, -\tfrac{1}{3}\right)\right] \tag{7.26}$$

eight colored weak singlets:

$$4\left(1, 3, -\tfrac{2}{3}\right) + 3\left(1, \bar{3}, +\tfrac{1}{3}\right) + \left(3, 1, \tfrac{2}{3}\right) \tag{7.27}$$

a weak singlet quix:

$$\left(\bar{6}, 1, -\tfrac{1}{3}\right) + \left(1, 6, \tfrac{1}{3}\right) \tag{7.28}$$

and three charged leptons and their neutrinos.

The quix plays an essential role for anomaly cancellation. The couplings of the axigluon are flavor-diagonal. However, the high quark is primarily a weak singlet, so that flavor violation in the $uc$ sector can be mediated by the $Z^0$. Such GIM violation is suppressed by $\phi^4$, where $\phi$ measures the mass mixing of the high quark with the $u$ and $c$ quarks. (Alternatively, GIM can be guaranteed by taking the quix to have $Q = -\frac{2}{3}$ and by assigning to the high quark the anomaly-canceling charge of $-\frac{13}{3}$. This possibility may be too ugly to be taken seriously.)

In the scalar meson sector, two $\phi(3, \bar{3}, 2)$ are required, one with charges $Q = 0, 1$ and one with charges $Q = 0, -1$. In addition, there must be a $\phi(1, 1, 2)$ to provide lepton masses and a $\phi(3, \bar{3}, 1)$ to give mass to the odd up quark. The quix can obtain its mass from a neutral $\phi(6, \bar{6})$ of scalar mesons, or through one-loop diagrams involving VEVless $\phi(3, 1, 2)$ and $\phi(3, 1, 1)$ scalar mesons. The latter possibility is more attractive since it presents a possible mechanism for quix decay.

**Mark III.** This is a four-family model with a quix involving four colored weak doublets:

$$4\left[\left(3, 1, \tfrac{2}{3}\right) + \left(3, 1, -\tfrac{1}{3}\right)\right] \tag{7.29}$$

eight colored weak singlets:

$$4\left(1, \bar{3}, -\tfrac{2}{3}\right) + 3\left(1, \bar{3}, +\tfrac{1}{3}\right) + \left(\bar{3}, 1, \tfrac{1}{3}\right) \tag{7.30}$$

a weak singlet quix

$$\left(\bar{6}, 1, -\tfrac{1}{3}\right) + \left(1, 6, \tfrac{1}{3}\right) \tag{7.31}$$

and four charged leptons and their neutrinos.

Couplings of the axigluon yield flavor-changing effects in the $dsb$ sector since one of the $Q = -\frac{1}{3}$ quarks (presumably low) is treated differently from the others. It is suppressed by $\phi^4$, where $\phi$ is a measure of the mass mixing of low with its lighter cousins. The scalar mesons needed are the same as for Mark II except that a $\phi(3, \bar{3}, 1)$ multiplet is not required.

**Mark IV.** This model involves fermions that transform under both SU(3) factors. There are four colored weak doublets:

$$4\left[\left(3, 1, \tfrac{2}{3}\right) + \left(3, 1, -\tfrac{1}{3}\right)\right] \tag{7.32}$$

eight colored weak singlets:

$$4\left(1, \bar{3}, \tfrac{1}{3}\right) + 2\left(1, \bar{3}, -\tfrac{2}{3}\right) + 2\left(\bar{3}, 1, -\tfrac{2}{3}\right) \tag{7.33}$$

two dichromatic fermion multiplets:

$$2(\bar{3}, 3, 0) \tag{7.34}$$

and four charged leptons and their neutrinos.

The axigluon can produce flavor violation in the $uc$ sector. The amplitude for $uc \leftrightarrow \bar{c}u$ mixing is suppressed by $\phi^4$ where $\phi$ is a measure of mass mixing between heavy and light $Q = \frac{2}{3}$ quarks.

Two scalar $\phi(3, \bar{3}, 2)$ multiplets with different charges and a $\phi(1, 1, 2)$ suffice to give mass to all ordinary quarks and leptons. A $\phi(3, \bar{3}, 1)$ which develops a VEV will give masses to the exotic fermions. In particular, the color singlet in each $(3, \bar{3})$ of fermions obtains just *twice* the mass as the corresponding color octet or quait. Additional scalar multiplets [such as a $\phi(3, 1, 1)$] will permit the eventual decay of the quait, should this be necessary or desirable.

**Mark V.** This model is for the hard-core GIM addict. It involves five standard families and two species of exotic colored fermions. There are five colored weak doublets:

$$5\left[\left(3, 1, \tfrac{2}{3}\right) + \left(3, 1, -\tfrac{1}{3}\right)\right] \tag{7.35}$$

ten colored weak singlets;

$$5\left(1, \bar{3}, -\tfrac{2}{3}\right) + 5\left(1, \bar{3}, \tfrac{1}{3}\right) \tag{7.36}$$

a quix:

$$\left(\bar{6}, 1, -\tfrac{1}{3}\right) + \left(1, 6, \tfrac{1}{3}\right) \tag{7.37}$$

a dichromatic fermion multiplet:

$$(\bar{3}, 3, 0) \tag{7.38}$$

and five charged leptons and their neutrinos.

For this model, GIM is sacrosanct for all Yukawa couplings of the axigluons, the $Z^0$, and the scalar mesons of the Higgs sector, which must include just about all the multiplets mentioned in Marks I through IV. It is possible [10] to make a unified version of Mark V, using the group $SU(4)^6$.

In chiral color many new particles are predicted. The axigluon affects the calculations of upsilon decay [11] and the radiative corrections to $e^+e^-$ annihilation [12]. In particular, the axigluon should be visible in di-jet distributions at hadron colliders as well as in a Jacobian peak for single jet cross sections [13]. At present the lower bound [14, 15] on the axigluon mass is about 1 TeV. This implies that the scale of breaking of chiral color must be higher than the weak scale.

## 7.5 Three Families and the 331 Model

The motivation for the 331 model lies in the explanation of three families. It testability arises from the prediction of new particles, particularly the *bilepton*.

In the SM each family separately cancels the triangle anomaly. A possible reason for three families is that in an extension of the SM, the extended families are anomalous but there is interfamily cancellation.

The three families must enter asymmetrically to set up the $(+1\ +1\ -2)$ type of anomaly cancellation. If we assume first that two families are treated symmetrically (sequentially), then the $-2$ may be expected to arise from

$$\left(\frac{Q(u)}{Q(d)}\right) = -2 \tag{7.39}$$

This is how it happens in the 331 model.

Take the gauge group

$$SU(3)_C \times SU(3)_L \times U(1)_X \tag{7.40}$$

and from it the standard $SU(2)_L$ is contained in $SU(3)_L$ while $U(1)_Y$ is contained in both $SU(3)_L$ and $U(1)_X$.

The first family is assigned:

$$\begin{pmatrix} u \\ d \\ D \end{pmatrix}_L \quad \bar{u}_L \bar{d}_L \bar{D}_L \tag{7.41}$$

which involves a $3_L$ of $S(3)_L$ with $X = -\frac{1}{3}$ (in general, the $X$ charge equals the electric charge of the central component of any $SU(3)_L$ triplet) and three $SU(3)_L$ singlets.

Similarly, the second family is assigned:

$$\begin{pmatrix} c \\ s \\ S \end{pmatrix}_L \quad \bar{c}_L s_L S_L \tag{7.42}$$

where again the $3_L$ has $X = -\frac{1}{3}$.

The third family is assigned differently from the first and second:

$$\begin{pmatrix} T \\ t \\ b \end{pmatrix}_L \quad \bar{T}_L \bar{t}_L \bar{b}_L \tag{7.43}$$

Here the nontrivial $SU(3)_L$ representation is a $\bar{3}_L$ with $X = +\frac{2}{3}$.

In this 331 model, the leptons are treated more democratically, being assigned to three $\bar{3}_L$'s of $SU(3)_L$, all with $X = 0$:

$$\begin{pmatrix} e^+ \\ \nu_e \\ e^- \end{pmatrix} \begin{pmatrix} \mu^+ \\ \nu_\mu \\ \mu^- \end{pmatrix} \begin{pmatrix} \tau^+ \\ \nu_\tau \\ \tau^- \end{pmatrix} \tag{7.44}$$

In this arrangement of quarks and leptons, all anomalies cancel. Nontrivial inter-family cancellations take place for $(3_L)^3$, $(3_L)^2 X$, and $X^3$. The number of families must be a multiple of three, and indeed equal to three if one wishes to avoid the "superfamily" problem, created by having a nontrivial multiple of the three families.

The Higgs sector contains:

$$\text{Triplets:} \quad \phi^\alpha(X = +1) \quad \phi'^\alpha(X = 0) \quad \phi''^\alpha(X = -1) \tag{7.45}$$

$$\text{Sextet:} \quad S^{\alpha\beta}(X = 0) \tag{7.46}$$

The breaking of $SU(3)_L$ to $SU(2)_L$ gives rise to five massive gauge bosons by the Higgs mechanism. One is a $Z'$ and the other four fall into two doublets under $SU(2)_L$, the bileptons $(Y^{--}, Y^-)$ and their charge conjugates $(Y^{++}, Y^+)$.

The scale $U$ at which the symmetry $SU(3)_L \times U(1)_X \rightarrow SU(2)_L \times U(1)_X$ breaks has an upper bound for the following reason: The embedding needs $\sin^2\theta_W(M_Z) < \frac{1}{4}$. But $\sin^2\theta_W(M_Z) > 0.23$ and increases through 0.25 at

$\mu \sim 3$ TeV. So $U$ must be appreciably below this to avoid a Landau pole $g_X \to \infty$. This implies that $M(Y^{--}, Y^{-})$, $M(Q) < 1$ TeV.

The heavy exotic quarks can be sought in the same way as top quarks were. The bilepton can be seen in $e^+e^-$ scattering in the backward direction and, most strikingly, in the direct channel of $e^-e^-$ scattering, where it appears as a sharp peak in the cross section.

Bileptons were predicted earlier in an SU(15) grand unification scheme [16]. This unification has the motivation of avoiding proton decay and is testable by the prediction of weak-scale leptoquarks. However, the 331 model has more elegance, particularly with regard to anomaly cancellation.

The lower mass limit on the bilepton comes from polarized muon decay [17, 18] and from muonium–antimuonium conversion searches. At the time of writing (1999), the best limit comes from the latter experiment [19] and is $M(Y^-) > 850$ GeV.

## 7.6 Conformality Constraints

The gauge hierarchy problem is a *theory-generated* problem of the very small ratio $M(W)/M_{\text{GUT}} \simeq 10^{-14}$, which typically arises in grand unified models. Of course, if grand unification can be avoided, there is no hierarchy—the problem is nullified. This is the situation in the conformality approach discussed in this section.

Until the late 1990s it was believed that a gauge theory in $d =$ four space–time dimensions could be conformal only in the presence of supersymmetry. For example, $\mathcal{N} = 4$ supersymmetric SU($N$) Yang–Mills theory was shown in 1983 [20] to be conformal for all finite $N$. The idea [21] that an $\mathcal{N} = 0$ nonsupersymmetric gauge theory could be conformal, and include the standard model, followed the demonstration [22] of field-string duality. Although the field-string duality was derived only for the $N \to \infty$ limit, the fact that the conformality was already known to survive to finite $N$ in some (supersymmetric) cases led one to expect it to survive to finite $N$ in new examples, especially some without supersymmetry.

With such conformality as a guide to extending the standard model, the gauge couplings cease to run above the conformality scale (here assumed to be a few TeV) and there is no grand unification or gauge hierarchy problem. This is the motivation for conformality, together with the aesthetic appeal of the absence of infinite renormalization. The testability of conformality arises from the prediction of many new particles, both fermions and scalars, at the TeV scale and producible at the next generation of colliders.

Until 1997 there was no reason to believe that conformal fixed points exist in gauge theories without supersymmetry in four space–time dimensions. Then it was pointed out [22] that compactification of a 10-dimensional type IIB superstring on a manifold $(\text{AdS})_5 \times S^5$ is of special interest in this regard; here $(\text{AdS})_5$ is five-dimensional anti-DeSitter space and $S^5$ is a five-sphere. On the four-dimensional surface of $(\text{AdS})_5$ exists a gauge field theory with interesting properties. Its gauge group is $SU(N)$ and arises from the coalescence of $N$ D3-branes in the manifold. The isotropy $SO(6) \sim SU(4)$ of the $S^5$ becomes the $R$ symmetry of $\mathcal{N} = 4$ supersymmetric Yang–Mills theory. The isometry $SO(4, 2)$ of $(AdS)_5$ is related to the conformal symmetry of the gauge theory in four dimensions. As already mentioned, the conformal symmetry of $N = 4$ gauge theory was known for some time [20].

What is especially interesting is the ability to break supersymmetry by replacing the manifold by an orbifold, in particular replacing $S^5$ by $S^5/\Gamma$, where $\Gamma$ is a freely acting discrete symmetry. The result, easy to derive, is that the breaking of the $\mathcal{N} = 4$ supersymmetries depends on the embedding of $\Gamma$ in the $SU(4)$ isotropy of $S^5$: If $\Gamma \subset SU(2)$, there remains $\mathcal{N} = 2$; if $\Gamma \subset SU(3)$, it leaves $\mathcal{N} = 1$; and if $\Gamma \not\subset SU(3)$, the gauge theory has $\mathcal{N} = 0$. All such gauge theories, even without supersymmetry, possess equality of numbers of fermions and bosons, and this underlies their potential finiteness and, a stronger requirement, high-energy conformality.

Now we can attempt to identify the content of chiral fermions and scalars in such theories. Some details of the rules are presented in Refs. [21, 23–26].

The embedding of $\Gamma$ is most simply illustrated for an abelian group $Z_p$. The action on the coordinates $(X_1, X_2, X_3)$ of the three-dimensional complex space $\mathcal{C}_3$ in which is the $S^5$ can be written in terms of the three integers $a_i = (a_1, a_2, a_3)$ such that the action of $Z_p$ is

$$C_3 : (X_1, X_2, X_3) \xrightarrow{Z_p} \left(\alpha^{a_1} X_1, \alpha^{a_2} X_2, \alpha^{a_3} X_3\right) \tag{7.47}$$

with

$$\alpha = \exp\left(\frac{2\pi i}{p}\right) \tag{7.48}$$

The scalar multiplet is in the **6** of $SU(4)$ $R$ symmetry and is transformed by the $Z_p$ transformation:

$$\operatorname{diag}\left(\alpha^{a_1}, \alpha^{a_2}, \alpha^{a_3}, \alpha^{-a_1}, \alpha^{-a_2}, \alpha^{-a_3}\right) \tag{7.49}$$

together with the gauge transformation

$$\operatorname{diag}\left(\alpha^0, \alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5\right) \times \alpha^i \tag{7.50}$$

for the different $SU(N)_i$ of the gauge group $SU(N)^p$.

What will be relevant are states invariant under a combination of these two transformations. If $a_1 + a_2 + a_3 = 0 \pmod p$, the matrix

$$\begin{pmatrix} a_1 & & \\ & a_2 & \\ & & a_3 \end{pmatrix} \tag{7.51}$$

is in $SU(3)$ and hence $\mathcal{N} \geqslant 1$ is unbroken and this condition must therefore be avoided if we want $\mathcal{N} = 0$.

If we examine the **4** of SU(4), we find that matter which is invariant under the combination of the $Z_p$ and an $SU(N)^p$ gauge transformation can be deduced similarly. It is worth defining the spinor **4** explicitly by $A_q = (A_1, A_2, A_3, A_4)$, with the $A_q$, like the $a_i$, defined only mod $p$. Explicitly, we may define $a_1 = A_1 + A_2$, $a_2 = A_2 + A_3$, $a_3 = A_3 + A_1$, and $A_4 = -(A_1 + A_2 + A_3)$. In other words, $A_1 = \frac{1}{2}(a_1 - a_2 + a_3)$, $A_2 = \frac{1}{2}(a_1 + a_2 - a_3)$, $A_3 = \frac{1}{2}(-a_1 + a_2 + a_3)$, and $A_4 = -\frac{1}{2}(a_1 + a_2 + a_3)$. To leave no unbroken supersymmetry we must obviously require that all $A_q$ are nonvanishing. In terms of the $a_i$ the condition that we shall impose is

$$\sum_{i=1}^{i=3} \pm(a_i) \neq 0 \pmod p \tag{7.52}$$

The $Z_p$ group identifies $p$ points in $C_3$. The $N$ converging D-branes approach all $p$ such points giving a gauge group with $p$ factors:

$$SU(N) \times SU(N) \times SU(N) \times \cdots \times SU(N) \tag{7.53}$$

The matter that survives is invariant under a product of a gauge transformation and a $Z_p$ transformation.

For the covering gauge group $SU(pN)$, the transformation is

$$\left(1, 1, \ldots, 1; \alpha, \alpha, \ldots, \alpha; \alpha^2, \alpha^2, \ldots, \alpha^2; \ldots; \alpha^{p-1}, \alpha^{p-1}, \ldots, \alpha^{p-1}\right) \tag{7.54}$$

with each entry occurring $N$ times.

Under the $Z_p$ transformation for the scalar fields, the **6** of SU(4), the transformation is

$$\sim X \quad \Rightarrow \quad \left(\alpha^{a_1}, \alpha^{a_2}, \alpha^{a_3}\right) \tag{7.55}$$

The result can conveniently by summarized by a *quiver diagram* [27]. One draws $p$ points and for each $a_k$ one draws a nondirected arrow between all modes $i$ and $i + a_k$. Each arrow denotes a bifundamental representation such that the resultant scalar representation is

$$\sum_{k=1}^{k=3} \sum_{i=1}^{i=p} (N_i, \bar{N}_{i \pm a_k}) \tag{7.56}$$

If $a_k = 0$, the bifundamental is to be reinterpreted as an adjoint representation plus a singlet representation.

For the chiral fermions one must construct the spinor **4** of SU(4). The components are the $A_q$ given above. The resultant fermion representation follows from a different quiver diagram. One draws $p$ points and connects with a *directed* arrow the node $i$ to the node $i + A_q$. The fermion representation is then

$$\sum_{q=1}^{q=4} \sum_{i=1}^{i=p} (N_i, \bar{N}_{i+A_q}) \tag{7.57}$$

Since all $A_q \neq 0$, there are no adjoint representations for fermions. This completes the matter representation of $SU(N)^p$.

The conformality approach is that the gauge particles and the quarks and leptons, together with some yet unseen degrees of freedom, may combine to give a quantum field theory with nontrivial realization of conformal invariance. In such a scenario the fact that there are no large mass corrections follows by the condition of conformal invariance. In other words, 't Hooft's naturalness condition is satisfied: namely, in the absence of masses there is an enhanced symmetry that is conformal invariance. We thus imagine the actual theory to be given by an action

$$S = S_0 + \int d^4x \, \alpha_i O_i \tag{7.58}$$

where $S_0$ is the Lagrangian for the conformal field theory in question and the $O_i$ are certain operators of dimension less than four, breaking conformal invariance softly. The $\alpha_i$ represent the mass parameters. Their mass dimension is $4 - \Delta_i$, where $\Delta_i$ is the dimension of the field $O_i$ at the conformal point. Note that the breaking should be soft, in order for the idea of conformal invariance to be relevant for solving the hierarchy problem and the cosmological constant problem. With conformal invariance the vacuum energy can have only one value, zero, and sufficiently soft breaking should preserve this value. This requires that the operators $O_i$ have dimension less than four.

Let $M$ denote the mass scale determined by the parameters $\alpha_i$. This is the scale at which the conformal invariance is broken. In other words, for energies $E \gg M$ the coupling will not run while they start running for $E < M$. We will assume that $M$ is sufficiently near the TeV scale to solve the hierarchy problem using conformal invariance.

The details of the conformal symmetry breaking are important to study further because (1) the prediction of dimensionless ratios in the standard model, such as mass ratios and mixing angles, are predicted before conformal breaking by purely group theoretical reasoning, since there is no flexibility in the Yukawa and quartic Higgs couplings, but then depend on the specific pattern assumed for conformal symmetry breaking, and (2) the cosmological constant is vanishing before conformal symmetry breaking and may remain so if the symmetry breaking is sufficiently mild.

We would like to discuss how the $SU(3) \times SU(2) \times U(1)$ standard model can be embedded in the conformal theories under discussion. In other words, we consider some embedding

$$SU(3) \times SU(2) \times U(1) \subset \bigotimes_i SU(Ndi) \tag{7.59}$$

in the set of conformal theories discussed in Section 7.3. Each gauge group of the standard model may lie in a single $SU(Nd_i)$ group or in some diagonal subgroup of a number of $SU(Nd_i)$ gauge groups in the conformal theory. The first fact to note, independent of the embeddings, is that the matter representations we will get in this way are severely restricted. This is because in the conformal theories we have only bifundamental fields (including adjoint fields), and thus any embedding of the standard model in the conformal theories under discussion will result in matter in bifundamentals (including adjoints), and no other representation. For example, we cannot have a matter field transforming according to representation of the form (**8**,**2**) of $SU(3) \times SU(2)$. That we can have only fundamental fields or bifundamental fields is a strong restriction on the matter content of the standard model, which in fact is satisfied, and we take it as a check (or evidence!) for the conformal approach to phenomenology. The rigidity of conformal theory in this regard can be compared to other approaches, where typically we can have various kinds of representations.

Another fact to note is that there are no U(1) factors in the conformal theories [having charged U(1) fields is in conflict with conformality], and in particular the existence of quantization of hypercharge is automatic in our setup, as the U(1) has to be embedded in some product of SU groups. This is the conformal version of the analogous statement in the standard scenarios to unification, such as SU(5) GUT.

This approach is examined in more detail in Chapter 8.

## 7.7 Summary

The standard model, although exceptionally successful in comparison with all high-energy data up to energies of 100 GeV, leaves several very fundamental questions unanswered. We have illustrated model building beyond the standard model by examples that are motivated by explanation of certain of these puzzling features.

The left–right model leads to the restoration of parity symmetry at high energy. It also leads to the gauging of the difference $(B - L)$ rather than the more obscure weak hypercharge. Tests of the left–right model include neutrino masses and the occurrence of various $B$ and $L$ violating processes.

In chiral color theory, the manner in which the color and electroweak interactions are treated so differently in the standard model is addressed. In this theory the color interaction, like the electromagnetic interaction, arises from spontaneous breaking of a large gauge group; this leads to the prediction of an octet of massive axigluons.

One of the most striking unexplained features of the standard model is surely the replication, three times, of the fermion families. In the 331 model, this arises from the cancellation of chiral anomalies between the families. One prediction of this model is the existence of bileptonic gauge bosons.

Finally, in the conformality approach, which is well motivated by the duality between string theory and gauge field theory, the standard model is assumed to simplify and become conformally invariant at multi-TeV energy. This leads to the prediction of new particles in that energy region, as will be discussed in Chapter 8.

Of course, these are only illustrative examples and the real world may be something not in the list above. Nevertheless, it is a valuable exercise to make attempts at such model building because the next generation of high-energy colliders will bring data that may or may not be surprising, depending on how many ideas for model building have already been explored.

## References


1 S.L. Glashow, *Nucl. Phys.* 22, 579 (1961).
2 A. Salam, in *Elementary Particle Theory*, N. Svartholm, ed., Almqvist & Wiksell, Stockholm, 1968.
3 S. Weinberg, *Phys. Rev. Lett.* 19, 1264 (1967).
4 G. 't Hooft, *Nucl. Phys.* B35, 167 (1971).
5 *Proceedings of the 19th International Conferences on High Energy Physics*, Tokyo, 1978.
6 J.C. Pati and A. Salam, *Phys. Rev.* D10, 275 (1974).
7 R.N. Mohapatra and J.C. Pati, *Phys. Rev.* D11, 566 (1975).
8 R.N. Mohapatra and G. Senjanovic, *Phys. Rev.* D12, 1502 (1975).
9 P.H. Frampton and S.L. Glashow, *Phys. Lett.* B190, 157 (1987).
10 P.H. Frampton and S. L. Glashow, *Phys. Rev. Lett.* 58, 2168 (1987).
11 F. Cuypers and P.H. Frampton, *Phys. Rev. Lett.* 60, 1237 (1988).
12 F. Cuypers, A.F. Falk, and P.H. Frampton, *Phys. Lett.* B259, 173 (1991).
13 F. Abe et al. (CDF collaboration), *Phys. Rev. Lett.* 71, 2542 (1993).
14 F. Abe et al. (CDF collaboration), *Phys. Rev.* D55, 6263 (1997).
15 M.A. Doncheski and R.W. Robinett, *Phys. Rev.* D58, 097702 (1998).
16 P.H. Frampton and B.H. Lee, *Phys. Rev. Lett.* 64, 619 (1990).
17 E.D. Carlson and P.H. Frampton, *Phys. Lett.* 64, 619 (1992).
18 A. Jodidio et al., *Phys. Rev.* D34, 1967 (1986).
19 L. Willmann et al., *Phys. Rev. Lett.* 82, 49 (1999).
20 S. Mandelstam, *Nucl. Phys.* B213, 149 (1983).
21 P.H. Frampton, *Phys. Rev.* D60, 041901 (1999).
22 J. Maldacena, *Adv. Theor. Math. Phys.* 2, 231 (1998).
23 P.H. Frampton and W.F. Shively, *Phys. Lett.* B454, 49 (1999).
24 P.H. Frampton and C. Vafa, hep-th/9903226.
25 P.H. Frampton, *Phys. Rev.* D60, 085004 (1999).
26 P.H. Frampton, *Phys. Rev.* D60, 121901 (1999).
27 M. Douglas and G. Moore, hep-m/9603167.

# 8
# Conformality

## 8.1
## Introduction

By now we have seen that the standard model of particle phenomenology is a gauge field theory based on the gauge group $SU_C(3) \times SU_L(2) \times U_Y(1)$ and with three families of quarks and leptons. The electroweak symmetry $SU_L(2) \times U_Y(1)$ is spontaneously broken by a Higgs mechanism to the electromagnetic symmetry $U_{EM}(1)$, leaving one Higgs boson. This model has successfully explained all experimental data (with the exception of neutrino masses which can be accommodated by extension). At higher energy than yet explored, the proliferation of parameters (19 without neutrino mass, 28 with) strongly suggests new physics beyond the standard model.

In this chapter, we discuss such an extension based on four-dimensional conformal invariance at high energy and inspired initially by the duality between gauge theory and superstring theory.

Such conformality model building is a less explored but equally motivated alternative to other directions of model building such as extra dimensions or supersymmetry.

Particle phenomenology is in an especially exciting time, mainly because of the anticipated data in a new energy regime expected from the large hadron collider (LHC), to be commissioned at the CERN Laboratory in 2007. This new data is long overdue. The superconducting supercollider (SSC) would have provided such data long ago were it not for its political demise in 1993.

Except for the remarkable experimental data concerning neutrino masses and mixings which has been obtained since 1998, particle physics has been data starved for the last 30 years. The standard model invented in the 1960s and 1970s has been confirmed and reconfirmed. Consequently, theory has ventured into speculative areas such as string theory, extra dimensions, and supersymmetry. While these ideas are of great interest and theoretically consistent there is no *direct* evidence from experiment for them. Here we describe a more recent, post 1998, direction known as conformality. First, to set the stage, we shall discuss why the conformality approach which is, in our opinion, competitive with the other three approaches, remained unstudied for the past 20 years up to 1998.

A principal motivation underlying model building, beyond the standard model, over the last 30 years has been the *hierarchy problem*, which is a special case of *naturalness*. This idea stems from Wilson [1] in the late 1960s. The definition of naturalness is that a theory should not contain any unexplained very large (or very small in the inverse) dimensionless numbers. The adjustment needed to achieve such naturalness violating numbers is called *fine tuning*. The naturalness situation can be especially acute in gauge field theories because even after fine tuning at tree level, i.e., the classical Lagrangian, the fine tuning may need to be repeated an infinite number of times order-by-order in the loop expansion during the renormalization process. While such a theory can be internally consistent it violates naturalness. Thus naturalness is not only an aesthetic criterion but one which the vast majority of the community feel must be imposed on any acceptable extension of the standard model; ironically, one exception is Wilson himself [2].

When the standard model of Glashow [3] was rendered renormalizable by appending the Higgs mechanism [4, 5] it was soon realized that it fell into trouble with naturalness, specifically through the hierarchy problem. In particular, the scalar propagator has quadratically divergent radiative corrections which imply that a bare Higgs mass $m_H^2$ will be corrected by an amount $\Lambda^2/m_H^2$, where $\Lambda$ is the cut off scale corresponding to new physics. Unlike logarithmic divergences, which can be absorbed in the usual renormalization process, the quadratic divergences create an unacceptable fine tuning: for example, if the cut off is at the conventional grand unification scale $\Lambda \sim 10^{16}$ GeV and $m_H \sim 100$ GeV, we are confronted with a preposterous degree of fine tuning to one part in $10^{28}$.

As already noted, this hierarchy problem was stated most forcefully by Wilson who said, in private discussions, that scalar fields are forbidden in gauge field theories. Between the late 1960s and 1974, it was widely recognized that the scalar fields of the standard model created this serious hierarchy problem but no one knew what to do about it.

The next big progress to the hierarchy problem came in 1974 with the invention [6] of supersymmetry. This led to the minimally supersymmetric standard model (MSSM) which elegantly answered Wilson's objection since quadratic divergences are cancelled between bosons and fermions, with only logarithmic divergences surviving. Further it was proved [7, 8] that the MSSM and straightforward generalizations were the unique way to proceed. Not surprisingly, the MSSM immediately became overwhelmingly popular. It has been estimated [9] that there are 35,000 papers existing on supersymmetry, more than an average of 1000 papers per year since its inception. This approach continued to seem "unique" until 1998. Since the MSSM has over 100 free parameters, many possibilities needed to be investigated and exclusion plots constructed. During this period, two properties beyond naturalness rendered the MSSM even more appealing: an improvement in unification properties and a candidate for cosmological dark matter.

Before jumping to 1998, it is necessary to mention an unconnected development in 1983 which was the study of Yang–Mills theory with extended $\mathcal{N} = 4$

supersymmetry (the MSSM has $\mathcal{N} = 1$ supersymmetry). This remarkable theory, though phenomenologically quite unrealistic as it allows no chiral fermions and all matter fields are in adjoint representations, is finite [10–12] to all orders of perturbation theory and conformally invariant. Between 1983 and 1997, the relationship between the $\mathcal{N} = 4$ gauge theory and either string theory, also believed to be finite, or the standard model remained unclear.

The perspective changed in 1997–1998 initially through the insight of Maldacena [13], who showed a *duality* between $\mathcal{N} = 4$ gauge theory and the superstring in ten space–time dimensions. Further the $\mathcal{N} = 4$ supersymmetry can be broken by orbifolding down to $\mathcal{N} = 0$ models with no supersymmetry at all. It was conjectured [14] by one of the authors in 1998 that such nonsupersymmetric orbifolded models can be finite and conformally invariant, hence the name conformality.

Conformality models have been investigated far less completely than supersymmetric ones but it is already clear that supersymmetry is "not as unique" as previously believed. No-go theorems can have not only explicit assumptions which need to be violated to avoid the theorem but unconscious implicit assumptions which require further progress even to appreciate: in 1975 the implicit assumption was that the gauge group is simple, or if semisimple may be regarded as a product of theories each with a simple gauge group. Naturalness, by cancellation of quadratic divergences, accurate unification, and a dark matter candidate exist in conformality.

It becomes therefore a concern that the design of the LHC has been influenced by the requirement of testing the MSSM. The LHC merits an investment of theoretical work to check if the LHC is adequately designed to test conformality which now seems equally as likely as supersymmetry, although we fully expect the detectors ATLAS and CMS to be sufficiently all purpose to capture any physics beyond the standard model at the TeV scale.

## 8.2 Quiver Gauge Theories

Quiver gauge theories possess a gauge group which is generically a product of $U(N_i)$ factors with matter fields in bifundamental representations. They have been studied in the physics literature since the 1980s where they were used in composite model building. They have attracted much renewed attention because of their natural appearance in the duality between superstrings and gauge theories.

The best known such duality gives rise to a highly supersymmetric ($\mathcal{N} = 4$) gauge theory with a single $SU(N)$ gauge group with matter in adjoint representations. In this case one can drop with impunity the $U(1)$ of $U(N)$ because the matter fields are uncharged under it. In the quiver theories with less supersymmetry ($\mathcal{N} \leqslant 2$) it is usually necessary to keep such $U(1)$s.

Quiver gauge theories are tailor-made for particle physics model building. While an $SU(N)$ gauge theory is typically anomalous in for arbitrary choice of fermions, choosing the fermions to lie in a quiver insures anomaly cancellation. Furthermore, the fermions in a quiver arrange themselves in bifundamental representations of the product gauge group. This nicely coincides with the fact that all known fundamental fermions are in bifundamental, fundamental, or singlet representations of the gauge group. The study of quiver gauge theories goes back to the earliest days of gauge theories and the standard model. Other notable early examples are the Pati–Salam model and the trinification model. A vast literature exists on this subject, but we will concentrate on post AdS/CFT conjecture quiver gauge theory work.

Starting from $AdS_5 \times S^5$ we only have an $SU(N)$ $\mathcal{N} = 4$ supersymmetric gauge theory. In order to break SUSY and generate a quiver gauge theory there are several options open to us. Orbifolds [15–18], conifolds [19–23], and orientifolds [24–29] have all played a part in building quiver gauge theories. Since our focus is quiver gauge theories in general, but via orbifolding of $AdS_5 \times M^5$ in particular, we will not discuss the other options in detail but should point out that orbifolding from the eleven-dimensional M theory point of view has also an active area of research [30–35]. Furthermore, we are interested in orbifolds where the manifold $M^5$ is the five sphere $S^5$. There are other possible choices for $M^5$ of relevance to model building [36, 37] but we will not explore these here either. In building models from orbifolded $AdS_5 \times S^5$, it is often convenient to break the quiver gauge group to the trinification [38] group $SU(3)^3$ or to the Pati–Salam [39–41] group $SU(4) \times SU(2) \times SU(2)$, but there are again other possibilities, including more complicated intermediate groups like the quartification [42–47] symmetry $SU(3)^4$ that treats quarks and leptons on an equal footing.

It is important to note that although the duality with superstrings is a significant guide to such model building, and it is desirable to have a string dual to give more confidence in consistency, we shall focus on the gauge theory description in the approach to particle phenomenology, as there are perfectly good quiver gauge theories that have yet to be derived from string duality.

The simplest superstring-gauge duality arises from the compactification of a type-IIB superstring on the cleverly chosen manifold

$$AdS_5 \times S^5$$

which yields an $\mathcal{N} = 4$ supersymmetry which is an especially interesting gauge theory that has been intensively studied and possesses remarkable properties of finiteness and conformal invariance for all values of $N$ in its $SU(N)$ gauge group. By "conformality," we shall mean conformal invariance at high energy, also for finite $N$.

For phenomenological purposes, $\mathcal{N} = 4$ is too much supersymmetry. Fortunately, it is possible to break supersymmetries and hence approach more nearly the real world, with less or no supersymmetry in a conformality theory.

By factoring out a discrete (either Abelian or non-Abelian) group and composing an orbifold

$$S^5/\Gamma$$

one may break $\mathcal{N} = 4$ supersymmetry to $\mathcal{N} = 2$, 1, or 0. Of special interest is the $\mathcal{N} = 0$ case.

We may take an Abelian $\Gamma = Z_p$ (non-Abelian cases will also be considered in this review) which identifies $p$ points in a complex three-dimensional space $\mathcal{C}_3$.

The rules for breaking the $\mathcal{N} = 4$ supersymmetry are:

– If $\Gamma$ can be embedded in an SU(2) of the original SU(4) R-symmetry, then

$$\Gamma \subset \mathrm{SU}(2) \quad \Rightarrow \quad \mathcal{N} = 2$$

– If $\Gamma$ can be embedded in an SU(3) but not an SU(2) of the original SU(4) R-symmetry, then

$$\Gamma \subset \mathrm{SU}(3) \quad \Rightarrow \quad \mathcal{N} = 1$$

– If $\Gamma$ can be embedded in the SU(4) but not an SU(3) of the original SU(4) R-symmetry, then

$$\Gamma \subset \mathrm{SU}(4) \quad \Rightarrow \quad \mathcal{N} = 0$$

In fact to specify the embedding of $\Gamma = Z_p$ we need to identify three integers $(a_1, a_2, a_3)$:

$$\mathcal{C}_3 : (X_1, X_2, X_3) \xrightarrow{Z_p} \left(\alpha^{a_1} X_1, \alpha^{a_2} X_2, \alpha^{a_3} X_3\right)$$

with

$$\alpha = \exp\left(\frac{2\pi i}{p}\right)$$

The $Z_p$ discrete group identifies $p$ points in $\mathcal{C}_3$. The $N$ converging D3-branes meet on all $p$ copies, giving a gauge group: $\mathrm{U}(N) \times \mathrm{U}(N) \times \cdots \times \mathrm{U}(N)$, $p$ times. The matter (spin-1/2 and spin-0) which survives is invariant under a product of a gauge transformation and a $Z_p$ transformation.

There is a convenient diagrammatic way to find the result from a "quiver." One draws $p$ points and arrows for $a_1, a_2, a_3$.

An example for $Z_5$ and $a_i = (1, 3, 0)$ as shown in Fig. 8.1.

For a general case, the scalar representation contains the bifundamental scalars

$$\sum_{k=1}^{3} \sum_{i=1}^{p} (N_i, \overline{N}_{i \pm a_k})$$

Note that by definition a *bifundamental* representation transforms as a fundamental ($N_i$) under $\mathrm{U}(N)_i$ and simultaneously as an antifundamental ($\overline{N}_{i \pm a_k}$) under $\mathrm{U}(N)_{i \pm a_k}$.

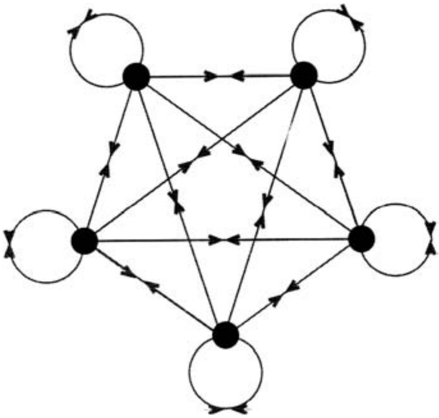


**Figure 8.1** Fermion quiver diagram.

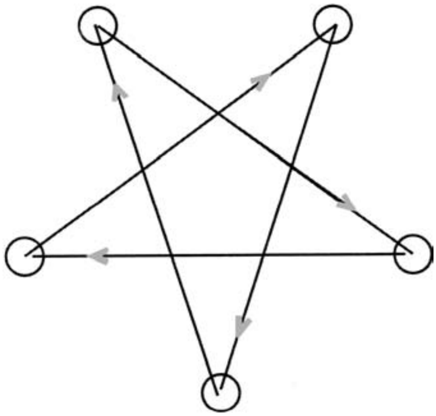


**Figure 8.2** Scalar quiver diagram.

For fermions, one must first construct the **4** of R-parity SU(4), isomorphic to the isometry SO(6) of the $S^5$. From the $a_k = (a_1, a_2, a_3)$ one constructs the 4-spinor $A_\mu = (A_1, A_2, A_3, A_4)$:

$$A_1 = \frac{1}{2}(a_1 + a_2 + a_3)$$

$$A_2 = \frac{1}{2}(a_1 - a_2 - a_3)$$

$$A_3 = \frac{1}{2}(-a_1 + a_2 - a_3)$$

$$A_4 = \frac{1}{2}(-a_1 - a_2 + a_3)$$

These transform as $\exp(\frac{2\pi i}{p} A_\mu)$ and the invariants may again be derived (by a different diagram). An example of a fermion quiver with $p = 5$ is shown above in Fig. 8.2.

Note that these lines are oriented, as is necessary to accommodate chiral fermions. Specifying the four $A_\mu$ is equivalent (there is a constraint that the four add to zero, mod $p$) to fixing the three $a_k$ and group theoretically is more fundamental.

In general, the fermion representation contains the bifundamentals

$$\sum_{\mu=1}^{4}\sum_{i=1}^{p}(N_i, \overline{N}_{i+A_\mu})$$

When one of the $A_\mu$s is zero, it signifies a degenerate case of a bifundamental comprised of adjoint and singlet representations of one $U(N)$.

To summarize the orbifold construction, first we select a discrete subgroup $\Gamma$ of the $SO(6) \sim SU(4)$ isometry of $S^5$ with which to form the orbifold $AdS_5 \times S^5/\Gamma$. As discussed above, the replacement of $S^5$ by $S^5/\Gamma$ reduces the supersymmetry to $\mathcal{N} = 0$, 1, or 2 from the initial $\mathcal{N} = 4$, depending on how $\Gamma$ is embedded in the isometry of $S^5$. The cases of interest here are $\mathcal{N} = 0$ and $\mathcal{N} = 1$ SUSY, where $\Gamma$ embeds irreducibly in the SU(4) isometry or in an SU(3) subgroup of the SU(4) isometry, respectively. It means to achieve $\mathcal{N} = 0$ we embed rep($\Gamma$) $\to$ **4** of SU(4) as $\mathbf{4} = (\mathbf{r})$, where $\mathbf{r}$ is a nontrivial four-dimensional representation of $\Gamma$; for $\mathcal{N} = 1$ we embed rep($\Gamma$) $\to$ **4** of SU(4) as $\mathbf{4} = (\mathbf{1}, \mathbf{r})$, where **1** is the trivial irreducible representation (irrep) of $\Gamma$ and $\mathbf{r}$ is a nontrivial three-dimensional representation of $\Gamma$.

## 8.3 Conformality Phenomenology

In attempting to go beyond the standard model, one outstanding issue is the hierarchy between GUT scale and weak scale which is 14 orders of magnitude. Why do these two very different scales exist? Also, how is this hierarchy of scales stabilized under quantum corrections? Supersymmetry answers the second question but not the first.

The idea is to approach hierarchy problem by conformality at a TeV scale. We will show how this is possible including explicit examples containing standard model states.

In some sense conformality provides more rigid constraints than supersymmetry. It can predict additional states at TeV scale, while there can be far fewer initial parameters in conformality models than in SUSY models. Conformality also provides a new approach to gauge coupling unification. It confronts naturalness and provides cancellation of quadratic divergences. The requirements of anomaly cancellations can lead to conformality of U(1) couplings.

There is a viable dark matter candidate, and proton decay can be consistent with experiment.

What is the physical intuition and picture underlying conformality? Consider going to an energy scale higher than the weak scale, for example at the TeV scale.

Quark and lepton masses, QCD and weak scales small compared to TeV scale. They may be approximated by zero. The theory is then classically conformally invariant though not at the quantum level because the standard model has nonvanishing renormalization group beta functions and anomalous dimensions. So this suggests that we add degrees of freedom to yield a gauge field theory with conformal invariance. There will be 't Hooft's naturalness since the zero mass limit increases symmetry to conformal symmetry.

We have no full understanding of how four-dimensional conformal symmetry can be broken spontaneously so breaking softly by relevant operators is a first step. The theory is assumed to be given by the action

$$S = S_0 + \int d^4x \alpha_i O_i \tag{8.1}$$

where $S_0$ is the action for the conformal theory and the $O_i$ are operators with dimension below four (i.e., relevant) which break conformal invariance softly.

The mass parameters $\alpha_i$ have mass dimension $4-\Delta_i$, where $\Delta_i$ is the dimension of $O_i$ at the conformal point.

Let $M$ be the scale set by the parameters $\alpha_i$ and hence the scale at which conformal invariance is broken. Then for $E \gg M$ the couplings will not run while they start running for $E < M$. To solve the hierarchy problem we assume $M$ is near the TeV scale.

Consider embedding the standard model gauge group according to

$$\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \subset \bigotimes_i \mathrm{U}(Nd_i)$$

Each gauge group of the SM can lie entirely in a $\mathrm{SU}(Nd_i)$ or in a diagonal subgroup of a number thereof.

Only bifundamentals (including adjoints) are possible. This implies no (8, 2), (3, 3), etc. A conformality restriction which is new and satisfied in Nature! The fact that the standard model has matter fields all of which can be accommodated as bifundamentals is experimental evidence for conformality.

No U(1) factor can be conformal in perturbation theory and so hypercharge is quantized through its incorporation in a non-Abelian gauge group. This is the "conformality" equivalent to the GUT charge quantization condition in, e.g., SU(5). It can explain the neutrality of the hydrogen atom. While these are postdictions, the predictions of the theory are new particles, perhaps at a low mass scale, filling out bifundamental representations of the gauge group that restore conformal invariance. The next section will begin our study of known quiver gauge theories from orbifolded $\mathrm{AdS}^5 \times S^5$.

**Table 8.1** All Abelian quiver theories with $\mathcal{N} = 0$ from $Z_2$ to $Z_7$

| | p | $A_m$ | $a_i$ | scal bfds | scal adjs | chir frms | SM |
|---|---|---|---|---|---|---|---|
| 1 | 2 | (1111) | (000) | 0 | 6 | No | No |
| 2 | 3 | (1122) | (001) | 2 | 4 | No | No |
| 3 | 4 | (2222) | (000) | 0 | 6 | No | No |
| 4 | 4 | (1133) | (002) | 2 | 4 | No | No |
| 5 | 4 | (1223) | (011) | 4 | 2 | No | No |
| 6 | 4 | (1111) | (222) | 6 | 0 | Yes | No |
| 7 | 5 | (1144) | (002) | 2 | 4 | No | No |
| 8 | 5 | (2233) | (001) | 2 | 4 | No | No |
| 9 | 5 | (1234) | (012) | 4 | 2 | No | No |
| 10 | 5 | (1112) | (222) | 6 | 0 | Yes | No |
| 11 | 5 | (2224) | (111) | 6 | 0 | Yes | No |
| 12 | 6 | (3333) | (000) | 0 | 6 | No | No |
| 13 | 6 | (2244) | (002) | 2 | 4 | No | No |
| 14 | 6 | (1155) | (002) | 2 | 4 | No | No |
| 15 | 6 | (1245) | (013) | 4 | 2 | No | No |
| 16 | 6 | (2334) | (011) | 4 | 2 | No | No |
| 17 | 6 | (1113) | (222) | 6 | 0 | Yes | No |
| 18 | 6 | (2235) | (112) | 6 | 0 | Yes | No |
| 19 | 6 | (1122) | (233) | 6 | 0 | Yes | No |
| 20 | 7 | (1166) | (002) | 2 | 4 | No | No |
| 21 | 7 | (3344) | (001) | 2 | 4 | No | No |
| 22 | 7 | (1256) | (013) | 4 | 2 | No | No |
| 23 | 7 | (1346) | (023) | 4 | 2 | No | No |
| 24 | 7 | (1355) | (113) | 6 | 0 | No | No |
| 25 | 7 | (1114) | (222) | 6 | 0 | Yes | No |
| 26 | 7 | (1222) | (333) | 6 | 0 | Yes | No |
| 27 | 7 | (2444) | (111) | 6 | 0 | Yes | No |
| 28 | 7 | (1123) | (223) | 6 | 0 | Yes | Yes |
| 29 | 7 | (1355) | (113) | 6 | 0 | Yes | Yes |
| 30 | 7 | (1445) | (113) | 6 | 0 | Yes | Yes |

## 8.4
## Tabulation of the Simplest Abelian Quivers

We consider the compactification of the type-IIB superstring on the orbifold $AdS_5 \times S^5/\Gamma$, where $\Gamma$ is an Abelian group $\Gamma = Z_p$ of the order $p$ with elements $\exp(2\pi i A/p)$, $0 \leqslant A \leqslant (p-1)$. Results for $p \leqslant 7$ are in Table 8.1.

The resultant quiver gauge theory has $\mathcal{N}$ residual supersymmetries with $\mathcal{N} = 2, 1, 0$ depending on the details of the embedding of $\Gamma$ in the SU(4) group which is the isotropy of the $S^5$. This embedding is specified by the four integers $A_m$, $1 \leqslant m \leqslant 4$ with

$$\sum_m A_m = 0 \bmod p$$

which characterize the transformation of the components of the defining representation of SU(4). We are here interested in the nonsupersymmetric case $\mathcal{N} = 0$ which occurs if and only if all four $A_m$ are nonvanishing.

## 8.5 Chiral Fermions

The gauge group is $U(N)^p$. The fermions are all in the bifundamental representations

$$\sum_{m=1}^{m=4} \sum_{j=1}^{j=p} (N_j, \overline{N}_{j+A_m}) \tag{8.2}$$

which are manifestly nonsupersymmetric because no fermions are in adjoint representations of the gauge group. Scalars appear in representations

$$\sum_{i=1}^{i=3} \sum_{j=1}^{i=p} (N_j, \overline{N}_{j\pm a_i}) \tag{8.3}$$

in which the six integers $(a_i, -a_i)$ characterize the transformation of the antisymmetric second-rank tensor representation of SU(4). The $a_i$ are given by $a_1 = (A_2 + A_3)$, $a_2 = (A_3 + A_1)$, and $a_3 = (A_1 + A_2)$.

It is possible for one or more of the $a_i$ to vanish in which case the corresponding scalar representation in the summation in Eq. (8.3) is to be interpreted as an adjoint representation of one particular $U(N)_j$. One may therefore have zero, two, four, or all six of the scalar representations, in Eq. (8.3), in such adjoints. One purpose of the present article is to investigate how the renormalization properties and occurrence of quadratic divergences depend on the distribution of scalars into bifundamental and adjoint representations.

Note that there is one model with all scalars in adjoints for each even value of $p$. For general even $p$ the embedding is $A_m = (\frac{p}{2}, \frac{p}{2}, \frac{p}{2}, \frac{p}{2})$. This series by itself forms the complete list of $\mathcal{N} = 0$ Abelian quivers with all scalars in adjoints.

To be of more phenomenological interest the model should contain chiral fermions. This requires that the embedding be complex: $A_m \not\equiv -A_m \pmod p$. It will now be shown that for the presence of chiral fermions all scalars must be in bifundamentals.

The proof of this assertion follows by assuming the contrary that there is at least one adjoint arising from, say, $a_1 = 0$. Therefore $A_3 = -A_2 \pmod{p}$. But then it follows from Eq. (8.2) that $A_1 = -A_4 \pmod{p}$. The fundamental representation of $SU(4)$ is thus real and fermions are nonchiral.[1)]

The converse also holds: If all $a_i \neq 0$ then there are chiral fermions. This follows since by assumption $A_1 \neq -A_2$, $A_1 \neq -A_3$, $A_1 \neq -A_4$. Therefore reality of the fundamental representation would require $A_1 \equiv -A_1$ hence, since $A_1 \neq 0$, $p$ is even and $A_1 \equiv \frac{p}{2}$; but then the other $A_m$ cannot combine to give only vector-like fermions.

It follows that

*In an $\mathcal{N} = 0$ quiver gauge theory, chiral fermions are possible if and only if all scalars are in bifundamental representations.*

We can prove a Mathematical Theorem:

*A pseudoreal* 4 of $SU(4)$ *cannot yield chiral fermions.*

In [48] it was proved that if the embedding in $SU(4)$ is such that the **4** is real: $\mathbf{4} = \mathbf{4}^*$, then the resultant fermions are always nonchiral. It was implied there that the converse holds, that if **4** is complex, $\mathbf{4} = \mathbf{4}^*$, then the resulting fermions are necessarily chiral. Actually for $\Gamma \subset SU(2)$ one encounters the intermediate possibility that the **4** is *pseudoreal*. In the present section we shall show that if **4** is pseudoreal then the resultant fermions are necessarily nonchiral. The converse now holds: if the **4** is neither real nor pseudoreal then the resultant fermions are chiral.

For $\Gamma \subset SU(2)$ it is important that the embedding be contained within the chain $\Gamma \subset SU(2) \subset SU(4)$ otherwise the embedding is not a consistent one. One way to see the inconsistency is to check the reality of the $\mathbf{6} = (\mathbf{4} \otimes \mathbf{4})_{\text{antisymmetric}}$. If $\mathbf{6} \neq \mathbf{6}^*$ then the embedding is clearly improper. To avoid this inconsistency it is sufficient to include in the **4** of $SU(4)$ only complete irreducible representations of $SU(2)$.

An explicit example will best illustrate this propriety constraint on embeddings. Let us consider $\Gamma = Q_6$, the dicyclic group of order $g = 12$. This group has six inequivalent irreducible representations: $1, 1', 1'', 1''', 2_1, 2_2$. The $1, 1', 2_1$ are real. The $1''$ and $1'''$ are a complex conjugate pair. The $2_2$ is pseudoreal. To embed $\Gamma = Q_6 \subset SU(4)$ we must choose from the special combinations which are complete irreducible representations of $SU(2)$ namely $1$, $2 = 2_2$, $3 = 1' + 2_1$ and $4 = 1'' + 1''' + 2_2$. In this way the embedding either makes the **4** of $SU(4)$ real, e.g., $4 = 1 + 1' + 2_1$ and the theorem of [48] applies, and nonchirality results, or the **4** is pseudoreal, e.g., $4 = 2_2 + 2_2$. In this case one can check that the embedding is consistent because $(\mathbf{4} \otimes \mathbf{4})_{\text{antisymmetric}}$ is real. But it is equally easy to check that the product of this pseudoreal **4** with the complete set of irreducible representations of $Q_6$ is again real and that the resultant fermions are nonchiral.

1) This is almost obvious but for a complete justification, see [48].

The lesson is:

*To obtain chiral fermions from compactification on* $\mathrm{AdS}_5 \times S_5/\Gamma$, *the embedding of* $\Gamma$ *in* SU(4) *must be such that the* 4 *of* SU(4) *is neither real nor pseudoreal.*

## 8.6 Model Building

The next step is to examine how the framework of quiver gauge theories can accommodate, as a sub theory, the standard model. This requires that the standard model gauge group and the three families of quarks and leptons with their correct quantum numbers be accommodated.

In such model building a stringent requirement is that the scalar sector, prescribed by the quiver construction, can by acquiring vacuum expection, values break the symmetry spontaneously to the desired sub theory. This is unlike most other model building where one *chooses* the representations for the scalars to accomplish this goal. Here the representations for the scalars are dictated by the orbifold construction.

One useful guideline in the symmetry breaking is that to break a semisimple $\mathrm{SU}(N)^n$ gauge group to its $\mathrm{SU}(N)$ diagonal subgroup requires at least $(n-1)$ bifundamental scalars, connected to one another such that each of the $n$ $\mathrm{SU}(N)$ factors is linked to all of the others: it is insufficient if the bifundamental scalars fragment into disconnected subsets.

We shall describe Abelian orbifolds [14, 49–52]. As will become clear Abelian orbifolds lead to accommodation of the standard model in unified groups $\mathrm{SU}(3)^n$ while non-Abelian orbifolds can lead naturally [53–56] to incorporation of the standard model in gauge groups such as $\mathrm{SU}(4) \times \mathrm{SU}(2) \times \mathrm{SU}(2)$ and generalizations.

We will now classify compactifications of the type-IIB superstring on $\mathrm{AdS}_5 \times S^5/\Gamma$, where $\Gamma$ is an Abelian group of order $n \leqslant 7$. Appropriate embedding of $\Gamma$ in the isometry of $S^5$ yields non-SUSY chiral models that can contain the standard model.

The first three-family $\mathrm{AdS}_5 \times S^5/\Gamma$ model of this type had $\mathcal{N} = 1$ SUSY and was based on a $\Gamma = Z_3$ orbifold [57], see also [58]. However, since then some of the most studied examples have been models without supersymmetry based on both Abelian [14, 49–52] and non-Abelian [53–56] orbifolds of $\mathrm{AdS}_5 \times S^5$. Recently, non-SUSY three family $Z_{12}$ orbifold models [59–61] have been shown to unify at a low scale ($\sim$4 TeV) and to have the promise of testability. One motivation for studying the non-SUSY case is that the need for supersymmetry is less clear as: (1) the hierarchy problem is absent or ameliorated, (2) the difficulties involved in breaking the remaining $\mathcal{N} = 1$ SUSY can be avoided if the orbifolding already results in completely broken SUSY, and (3) many of the effects of SUSY are still present in the theory, just hidden. For example, the Bose–Fermi state count matches, RG equations preserve vanishing $\beta$ functions to some number of loops, etc. Here we concentrate on Abelian orbifolds with and without supersymmetry, where the orbifolding group $\Gamma$ has the order $n = o(\Gamma) \leqslant 12$. We systematically study those cases with chiral matter. We find all chiral models for $n \leqslant 7$. Several of these contain the standard model (SM) with three families.

For $\mathcal{N} = 0$ the fermions are given by $\sum_i \mathbf{4} \otimes R_i$ and the scalars by $\sum_i \mathbf{6} \otimes R_i$, where the set $\{R_i\}$ runs over all the irreps of $\Gamma$. For $\Gamma$ Abelian, the irreps are all one-dimensional and as a consequence of the choice of $N$ in the $1/N$ expansion, the gauge group [62] is $SU(N)^n$. Chiral models require the $\mathbf{4}$ to be complex ($\mathbf{4} \neq \mathbf{4}^*$) while a proper embedding requires $\mathbf{6} = \mathbf{6}^*$, where $\mathbf{6} = (\mathbf{4} \otimes \mathbf{4})_{\text{antisymmetric}}$. (Mathematical consistency requires $\mathbf{6} = \mathbf{6}^*$, see [48].)

We choose $N = 3$ throughout. This means that most of our models will proceed to the SM through trinification. It is also possible to start with larger $N$, say $N = 4$ and proceed to the SM or MSSM via Pati–Salam models. The analysis is similar, so the $N = 3$ case is sufficient to gain an understanding of the techniques needed for model building, what choice of $N$ leads to the optimal model is still an open question.

If $SU_L(2)$ and $U_Y(1)$ are embedded in diagonal subgroups $SU^p(3)$ and $SU^q(3)$, respectively, of the initial $SU^n(3)$, the ratio $\frac{\alpha_2}{\alpha_Y}$ is $\frac{p}{q}$, leading to a calculable initial value of $\theta_W$ with, $\sin^2\theta_W = \frac{3}{3+5(\frac{p}{q})}$. The more standard approach is to break the initial $SU^n(3)$ to $SU_C(3) \otimes SU_L(3) \otimes SU_R(3)$, where $SU_L(3)$ and $SU_R(3)$ are embedded in diagonal subgroups $SU^p(3)$ and $SU^q(3)$ of the initial $SU^n(3)$. We then embed all of $SU_L(2)$ in $SU_L(3)$ but $\frac{1}{3}$ of $U_Y(1)$ in $SU_L(3)$ and the other $\frac{2}{3}$ in $SU_R(3)$. This modifies the $\sin^2\theta_W$ formula to: $\sin^2\theta_W = \frac{3}{3+5(\frac{\alpha_2}{\alpha_Y})} = \frac{3}{3+5(\frac{3p}{p+2q})}$, which coincides with the previous result when $p = q$. One should use the second (standard) embedding when calculating $\sin^2\theta_W$ for any of the models obtained below. Note, if $\Gamma = Z_n$ the initial $\mathcal{N} = 0$ orbifold model (before any symmetry breaking) is completely fixed (recall we always are taking $N = 3$) by the choice of $n$ and the embedding $\mathbf{4} = (\alpha^i, \alpha^j, \alpha^k, \alpha^l)$, so we define these models by $M^n_{ijkl}$. The conjugate models $M^n_{n-i,n-j,n-k,n-l}$ contain the same information, so we need not study them separately.

To get a feel for the constructions, we begin this section by studying the first few $\mathcal{N} = 0$ chiral $Z_n$ models. First, when $\mathcal{N} = 0$, the only allowed $\Gamma = Z_2$ orbifold where $\mathbf{4} = (\alpha, \alpha, \alpha, \alpha)$ and $Z_3$ orbifold where $\mathbf{4} = (\alpha, \alpha, \alpha^2, \alpha^2)$ have only real representations and therefore will not yield chiral models. Next, for $\Gamma = Z_4$ the choice $\mathbf{4} = (\alpha, \alpha, \alpha, \alpha)$ with $N = 3$ where $\alpha = e^{\frac{\pi i}{2}}$ (in what follows we will write $\alpha = e^{\frac{2\pi i}{n}}$ for the roots of unity that generate $Z_n$), yields an $SU(3)^4$ chiral model. In the scalar content of this model a VEV for say a $(3, 1, \bar{3}, 1)$ breaks the symmetry to $SU_D(3) \times SU_2(3) \times SU_4(3)$ but renders the model vector-like, and hence uninteresting, so we consider it no further. (We consider only VEVs that cause symmetry breaking of the type $SU(N) \times SU(N) \to SU_D(N)$. Other symmetry breaking patterns are possible, but for the sake of simplicity they will not be studied here. It is clear from this and previous remarks that there are many phenomenological avenues involving quiver gauge theories yet to be explored.) The only other choice of embedding is a nonpartition model with $\Gamma = Z_4$ is $\mathbf{4} = (\alpha, \alpha, \alpha, \alpha^3)$ but it leads to the same scalars with half the chiral fermions so we move on to $Z_5$.

There is one chiral model for $\Gamma = Z_5$ and it is fixed by choosing $\mathbf{4} = (\alpha, \alpha, \alpha, \alpha^2)$, leading to $\mathbf{6} = (\alpha^2, \alpha^2, \alpha^2, \alpha^3, \alpha^3, \alpha^3)$ with real scalars. It is straightforward to write down the particle content of this $M^5_{1112}$ model. The best one can do toward the construction of the standard model is to give a VEV to a $(3, 1, \bar{3}, 1, 1)$ to break the $SU^5(3)$ symmetry to $SU_D(3) \times SU_2(3) \times SU_4(3) \times SU_5(3)$. Now a VEV for $(1, 3, \bar{3}, 1)$ completes the breaking to $SU^3(3)$, but the only remaining chiral fermions are $2[(3, \bar{3}, 1) + (1, 3, \bar{3}) + (\bar{3}, 1, 3)]$ which contains only two families.

Moving on to $\Gamma = Z_6$ we find two models where, as with the previous $Z_5$ model, the $\mathbf{4}$ is arranged so that $\mathbf{4} = (\alpha^i, \alpha^j, \alpha^k, \alpha^l)$ with $i + j + k + l = n$. These have $\mathbf{4} = (\alpha, \alpha, \alpha, \alpha^3)$ and $\mathbf{4} = (\alpha, \alpha, \alpha^2, \alpha^2)$ and were defined as partition models in [58] when $i$ was equal to zero. Here we generalize and call all models satisfying $i + j + k + l = n$ partition models.

A new class of models appears; these are the double partition models. They have $i + j + k + l = 2n$ and none are equivalent to single partition models (if we require that $i$, $j$, $k$, and $l$ are all positive integers) with $i + j + k + l = n$. The $\mathcal{N} = 1$ nonpartition models have been classified [56], and we find 11 $\mathcal{N} = 0$ examples in Table 4. While they have a self-conjugate $\mathbf{6}$, this is only a necessary condition that may be insufficient to insure the construction of viable string theory based models [61]. However, the $\mathcal{N} = 0$ nonpartition models may still be interesting phenomenologically and as a testing ground for models with the potential of broken conformal invariance.

Now let us return to $\Gamma = Z_6$ where the partition models of interest are: (1) $\mathbf{4} = (\alpha, \alpha, \alpha^2, \alpha^2)$ where one easily sees that VEVs for $(3, 1, \bar{3}, 1, 1, 1)$ and then $(1, 3, \bar{3}, 1, 1)$ lead to at most two families, while other SSB routes lead to equal or less chirality; and (2) $\mathbf{4} = (\alpha, \alpha, \alpha, \alpha^3)$ where VEVs for $(3, 1, \bar{3}, 1, 1, 1)$ followed by a VEV for $(1, 3, \bar{3}, 1, 1)$ leads to an $SU(3)^4$ model containing fermions $2[(3, \bar{3}, 1, 1) + (1, 3, \bar{3}, 1) + (1, 1, 3, \bar{3}) + (\bar{3}, 1, 1, 3)]$. However, there are insufficient scalars to complete the symmetry breaking to the standard model. In fact, one cannot even achieve the trinification spectrum.

We move on to $Z_7$, where there are three partition models: (1) for $\mathbf{4} = (\alpha, \alpha^2, \alpha^2, \alpha^2)$, we find no SSB pathway to the SM. There are paths to an SM with less than three families, e.g., VEVs for $(3, 1, 1, \bar{3}, 1, 1, 1)$, $(1, 3, 1, \bar{3}, 1, 1)$, $(3, \bar{3}, 1, 1, 1)$, and $(1, 3, \bar{3}, 1)$ lead to one family at the $SU^3(3)$ level; (2) for $\mathbf{4} = (\alpha, \alpha, \alpha, \alpha^4)$, again we find only paths to family-deficient standard models. An example is where we have VEVs for $(3, 1, \bar{3}, 1, 1, 1, 1)$, $(1, 3, \bar{3}, 1, 1, 1)$, $(3, 1, \bar{3}, 1, 1)$, and $(1, 3, \bar{3}, 1)$, which lead to a two-family $SU^3(3)$ model; (3) finally, $\mathbf{4} = (\alpha, \alpha, \alpha^2, \alpha^3)$ is the model discovered in [51], where VEVs to $(1, 3, 1, \bar{3}, 1, 1, 1)$, $(1, 1, 3, \bar{3}, 1, 1)$, $(1, 1, 3, \bar{3}, 1)$ and $(1, 1, 3, \bar{3})$ lead to a three-family model with the correct Weinberg angle at the $Z$-pole, $\sin^2\theta_W = 3/13$.

## 8.7 Summary

It has been established that conformality can provide (i) naturalness without one-loop quadratic divergence for the scalar mass [66] and anomaly cancellation [67]; (ii) precise unification of the coupling constants [59, 60], and (iii) a viable dark matter candidate [65]. It remains for experiment to check that quiver gauge theories with gauge group $U(3)^3$ or $U(3)^n$ with $n \geqslant 4$ are actually employed by Nature.

For completeness, we should note that possible problems with $\mathcal{N} = 0$ orbifolds have been pointed out both in one-loop calculations in field theory [63] and from studies of tachyonic instability in the ancestral string theory [64].

One technical point worth to be mention is that while the $U(1)$ anomalies are cancelled in string theory, in the gauge theory there is a different description [67] of such cancellation which has the advantage of suggesting how $U(1)$ gauge couplings may be conformally invariant at high energies. This is important because in string theory, except for special linear combinations, all such $U(1)$ factors acquire mass by the Stückelberg mechanism while in the gauge field theory the cancellation of quadratic divergences and solution of the hierarchy problem for $\mathcal{N} = 0$ require they be instead at the teravolt scale.

We have described how phenomenology of conformality has striking resonances with the standard model, as we have described optimistically as experimental evidence in its favor. It has been shown elsewhere [59] how 4 TeV unification predicts three families and new particles around 4 TeV accessible to experiment (LHC).

It is encouraging that the scalar propagator in these theories has no quadratic divergence if and only if there are chiral fermions. Anomaly cancellation in the effective Lagrangian has been tied to the conformality of $U(1)$ gauge couplings.

A dark matter candidate (LCP = lightest conformality particle) [65] may be produced at the large hadron collider (LHC), then directly detected from the cosmos. Study of proton decay [68] leads to the conclusion that quark and lepton masses arise not from the Yukawa couplings of the standard model but from operators induced in breaking of the four-dimensional conformal invariance. This implies that the Higgs couplings differ from those usually assumed in the unadorned standard model. This is yet another prediction from conformality to be tested when the Higgs scalar couplings and decay products are examined at the LHC in the near future.

## References


**1** K.G. Wilson, *Phys. Rev.* D3, 1818 (1971).

**2** K.G. Wilson, *Nucl. Phys. Proc. Suppl.* 140, 3 (2005). hep-lat/0412043.

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

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

**5** A. Salam, in Svartholm: Elementary Particle Theory, *Proceedings of the Nobel Symposium held 1968 at Lerum*, Sweden, Stockholm, 1968, pp. 367–377.

**6** J. Wess and B. Zumino, *Nucl. Phys.* B70, 39 (1974).

7 R. Haag, J.T. Lopuszansi and M. Sohnius, *Nucl. Phys.* B88, 257 (1975).
8 S. Coleman and J.E. Mandula, *Phys. Rev.* 159, 1251 (1967).
9 P. Woit, *Not Even Wrong*, Basic Books, New York, 2006.
10 S. Mandelstam, *Nucl. Phys.* B213, 149 (1983).
11 P. Howe and K. Stelle, *Phys. Lett.* B137, 135 (1984).
12 L. Brink, Talk at Johns Hopkins Workshop on Current Problems in High Energy Particle Theory, Bad Honnef, Germany, 1983.
13 J.M. Maldacena, *Adv. Theor. Math. Phys.* 2, 231 (1998). hep-th/9711200.
14 P.H. Frampton, *Phys. Rev.* D60, 041901 (1999). hep-th/9812117.
15 Z. Kakushadze, *Phys. Lett.* B491, 317 (2000). hep-th/0008041.
16 L.J. Dixon, J.A. Harvey, C. Vafa and E. Witten, *Nucl. Phys.* B274, 285 (1986).
17 L.J. Dixon, J.A. Harvey, C. Vafa and E. Witten, *Nucl. Phys.* B261, 678 (1985).
18 Z. Kakushadze and S.H.H. Tye, *Phys. Lett.* B392, 335 (1997). hep-th/9609027.
19 A.M. Uranga, *JHEP* 9901, 022 (1999). hep-th/9811004.
20 K. Oh and R. Tatar, *JHEP* 9910, 031 (1999). hep-th/9906012.
21 I.R. Klebanov and M.J. Strassler, *JHEP* 0008, 052 (2000). hep-th/0007191.
22 A. Hebecker and M. Ratz, *Nucl. Phys.* B670, 3 (2003). hep-ph/0306049.
23 S. Benvenuti and A. Hanany, *JHEP* 0508, 024 (2005). hep-th/0502043.
24 Z. Kakushadze, *Nucl. Phys.* B544, 265 (1999). hep-th/9808048.
25 Z. Kakushadze, *Phys. Rev.* D59, 045007 (1999). hep-th/9806091.
26 A. Fayyazuddin and M. Spalinski, *Nucl. Phys.* B535, 219 (1998). hep-th/9805096.
27 Z. Kakushadze, *Phys. Rev.* D58, 106003 (1998). hep-th/9804184.
28 Z. Kakushadze, *Nucl. Phys.* B529, 157 (1998). hep-th/9803214.
29 Z. Kakushadze, *JHEP* 0110, 031 (2001). hep-th/0109054.
30 Z. Lalak, A. Lukas and B.A. Ovrut, *Phys. Lett.* B425, 59 (1998). hep-th/9709214.
31 M. Faux, D. Lust and B.A. Ovrut, *Nucl. Phys.* B554, 437 (1999). hep-th/9903028.
32 M. Faux, D. Lust and B.A. Ovrut, *Nucl. Phys.* B589, 269 (2000). hep-th/0005251.
33 M. Faux, D. Lust and B.A. Ovrut, *Int. J. Mod. Phys.* A18, 3273 (2003). hep-th/0010087.
34 M. Faux, D. Lust and B.A. Ovrut, *Int. J. Mod. Phys.* A18, 2995 (2003). hep-th/0011031.
35 C.F. Doran, M. Faux and B.A. Ovrut, *Adv. Theor. Math. Phys.* 6, 329 (2003). hep-th/0108078.
36 G. Aldazabal, L.E. Ibanez, F. Quevedo, and A.M. Uranga, *JHEP* 0008, 002 (2000). hep-th/0005067.
37 Y.H. He, hep-th/0408142.
38 S.L. Glashow, in *Proceedings of the Fifth Workshop on Grand Unification*, K. Kang, H. Fried, and P.H. Frampton, eds., World Scientific, Singapore, 1984, pp. 88–94.
39 J.C. Pati and A. Salam, *Phys. Rev.* D10, 275 (1974).
40 T. Blazek, S.F. King and J.K. Parry, *JHEP* 0305, 016 (2003). hep-ph/0303192.
41 J.B. Dent and T.W. Kephart, arXiv:0705.1995 [hep-ph].
42 K.S. Babu, E. Ma and S. Willenbrock, *Phys. Rev.* D69, 051301 (2004). hep-ph/0307380.
43 S.L. Chen and E. Ma, *Mod. Phys. Lett.* A19, 1267 (2004). hep-ph/0403105.
44 A. Demaria, C.I. Low, and R.R. Volkas, *Phys. Rev.* D72, 075007 (2005) [*Erratum ibid.* D73, 079902 (2006)]. hep-ph/0508160.
45 A. Demaria, C.I. Low, and R.R. Volkas, *Phys. Rev.* D74, 033005 (2006). hep-ph/0603152.
46 A. Demaria and K.L. McDonald, *Phys. Rev.* D75, 056006 (2007). hep-ph/0610346.
47 K.S. Babu, T.W. Kephart, and H. Paes, arXiv:0709.0765 [hep-ph].
48 P.H. Frampton and T.W. Kephart, *Int. J. Mod. Phys.* A19, 593 (2004). hep-th/0306207.
49 P.H. Frampton and W.F. Shively, *Phys. Lett.* B454, 49 (1999). hep-th/9902168.
50 P.H. Frampton and C. Vafa, hep-th/9903226.
51 P.H. Frampton, *Phys. Rev.* D60, 085004 (1999). hep-th/9905042.
52 P.H. Frampton, *Phys. Rev.* D60, 121901 (1999). hep-th/9907051.
53 P.H. Frampton and T.W. Kephart, *Phys. Lett.* B485, 403 (2000). hep-th/9912028.
54 P.H. Frampton, *J. Math. Phys.* 42, 2915 (2001). hep-th/0011165.

**55** P.H. Frampton and T.W. Kephart, *Phys. Rev.* D64, 086007 (2001). hep-th/0011186.

**56** P.H. Frampton, R.N. Mohapatra and S. Suh, *Phys. Lett.* B520, 331 (2001). hep-ph/0104211.

**57** S. Kachru and E. Silverstein, *Phys. Rev. Lett.* 80, 4855 (1998). hep-th/9802183.

**58** T.W. Kephart and H. Paes, *Phys. Lett.* B522, 315 (2001). hep-ph/0109111.

**59** P.H. Frampton, *Mod. Phys. Lett.* A18, 1377 (2003). hep-ph/0208044.

**60** P.H. Frampton, R.M. Rohm and T. Takahashi, *Phys. Lett.* B570, 67 (2003). hep-ph/0302074.

**61** P.H. Frampton and T.W. Kephart, *Phys. Lett.* B585, 24 (2004). hep-th/0306053.

**62** A. Lawrence, N. Nekrasov and C. Vafa, *Nucl. Phys.* B533, 199 (1998). hep-th/9803015.

**63** A. Adams and E. Silverstein, *Phys. Rev.* D64, 086001 (2001). hep-th/0103220.

**64** A. Adams, J. Polchinski and E. Silverstein, *JHEP* 0110, 029 (2001). hep-th/0108075.

**65** P.H. Frampton, *Mod. Phys. Lett.* A22, 921 (2007). astro-ph/0607391.

**66** X. Calmet, P.H. Frampton and R.M. Rohm, *Phys. Rev.* D72, 055033 (2005). hep-th/0412176.

**67** E. Di Napoli and P.H. Frampton, *Phys. Lett.* B638, 374 (2006). hep-th/0412176.

**68** P.H. Frampton, *Mod. Phys. Lett.* A22, 347 (2007). hep-ph/0610116.

# Index

***g***

## h



## i

**About the Second Edition:** The second edition of the successful textbook *Gauge Field Theories* by Paul Frampton has been updated from the original 1986 edition. The introductory material on gauge invariance and renormalization has been reorganized in a more logical manner. The fourth chapter on electroweak interactions has been rewritten to include the present status of precision electroweak data, and the discovery of the top quark. After the treatment of the renormalization group and of quantum chromodynamics, the final, seventh, chapter is a newly written one on model building. This book has been adopted for postgraduate courses by universities and institutions of higher learning around the world. The updated second edition has been long awaited and will be of interest to all students who aspire to undertake research in theoretical physics.

**About the Author:** Paul Frampton is the Louis D. Rubin, Jr. Professor of Physics at the University of North Carolina at Chapel Hill. His research interests in theoretical physics lie in particle phenomenology, string theory, and cosmology. He was born in Kidderminster, England and was educated in the University of Oxford which awarded him a BA degree with Double First Class honors in 1965, followed by a D. Phil in 1968 and an advanced D.Sc. degree in 1984. Frampton has published 250 scientific articles and two books: *Dual Resonance Models and Superstrings* (Reprinted) and *Gauge Field Theories* (First Edition). He is a Fellow of the American Physical Society, a Fellow of the British Institute of Physics, and a Fellow of the American Association for the Advancement of Science.

Professor Frampton has written important papers in both cosmology (the first calculation of vacuum decay) and string theory (first calculation of hexagon anomaly) and is also widely known for contributions (chiral color, 331-model, etc.) to particle phenomenology. He has supervised 30 PhD students and postdoctoral researchers.