RESOURCE LETTER

Roger H. Stuewer, *Editor*
School of Physics and Astronomy, 116 Church Street
University of Minnesota, Minneapolis, Minnesota 55455

This is one of a series of Resource Letters on different topics intended to guide college physicists, astronomers, and other scientists to some of the literature and other teaching aids that may help improve course contents in specified fields. No Resource Letter is meant to be exhaustive and complete; in time there may be more than one letter on some of the main subjects of interest. Comments on these materials as well as suggestions for future topics will be welcomed. Please send such communications to Professor Roger H. Stuewer, Editor, AAPT Resource Letters, School of Physics and Astronomy, 116 Church Street SE, University of Minnesota, Minneapolis, MN 55455.

Resource Letter: GI-1 Gauge invariance

T. P. Cheng
Department of Physics, University of Missouri, St. Louis, Missouri 63121[a] *and School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540*

Ling-Fong Li
Department of Physics, Carnegie–Mellon University, Pittsburgh, Pennsylvania 15213[a] *and School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540*

(Received 15 February 1988; accepted for publication 1 March 1988)

Gauge symmetry has become one of the most basic concepts in the theoretical framework for understanding fundamental interactions: The quantum field theories of the strong, weak, and electromagnetic interactions are all gauge invariant; general relativity can also be viewed as a classical gauge theory. This Resource Letter is intended for nonspecialists who wish to study gauge theories and their applications to elementary particle physics. About half of our discussion is devoted to the general properties of gauge-invariant quantum field theories and half to their applications, from QED, QCD, and the Weinberg–Salam model to grand unification. The letter E after an item indicates elementary level or material of general interest to persons becoming informed in the field. The letter I, for intermediate level, indicates material of somewhat more specialized nature; and the letter A indicates rather specialized or advanced material. An asterisk (*) indicates those articles to be included in an accompanying Reprint Book.

CONTENTS

I. INTRODUCTION

We begin with a brief discussion of the *gauge principle*, and its early history. Recall the familiar concept of gauge invariance in classical electrodynamics. The usual Maxwell equations are expressed in terms of the electromagnetic fields **E** and **B**:

$$\nabla\cdot\mathbf{B}=0,\quad \nabla\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0, \tag{1}$$

$$\nabla\cdot\mathbf{E}=\rho,\quad \nabla\times\mathbf{B}-\frac{\partial\mathbf{E}}{\partial t}=\mathbf{j}, \tag{2}$$

where ρ and **j** are, respectively, the charge and current densities. (We have adopted the Heaviside–Lorentz unit system and set the velocity of light $c=1$.) The homogeneous equations in Eq. (1) can be easily solved by introducing the electromagnetic potentials ϕ and **A**, which are related to the **E** and **B** fields by

$$\mathbf{B}=\nabla\times\mathbf{A},\quad \mathbf{E}=-\nabla\phi-\frac{\partial\mathbf{A}}{\partial t}. \tag{3}$$

The inhomogeneous equations in Eq. (2) govern the dynamics of ϕ and **A**. This simplifies the description of the electromagnetic field by reducing the dynamical variable from six to four. However, given a set of **E** and **B** fields, the corresponding potentials are not unique. Namely, **E** and **B** (hence, the Maxwell equations) are invariant under a *gauge transformation*

$$\phi\to\phi-\frac{\partial\theta}{\partial t},\quad \mathbf{A}\to\mathbf{A}+\nabla\theta, \tag{4}$$

where θ is an arbitrary space-time-dependent scalar function, the *gauge function.*

Classically, gauge invariance does not seem to be very profound. It merely represents the freedom in the choice of the electromagnetic potential for the description of electromagnetic phenomena. And we can describe the classical system completely in terms of the physically measurable **E** and **B** fields without any arbitrariness. However, when we go from classical to quantum systems, the situation changes. The quantum mechanics of electromagnetic interactions are most simply formulated in terms of the electromagnetic potentials ϕ and **A** rather than the **E** and **B** fields. Thus gauge invariance plays an essential role in the quantum description of electromagnetism. For example, the Schrödinger equation for an electron in the presence of an electromagnetic field is given by (with $\hbar = 1$):

$$\left(-\frac{1}{2m} (\nabla + ie\mathbf{A})^2 - e\phi \right)\psi = i\frac{\partial \psi}{\partial t}, \qquad (5)$$

as one can show, starting with the Lorentz force law, that the canonical conjugate momentum differs from the kinematic momentum by a factor of $e\mathbf{A}$. This equation is related to that for the free particle by the substitution:

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - ie\phi, \quad \nabla \to \nabla + ie\mathbf{A}. \qquad (6)$$

This is called the *principle of minimal coupling*, which introduces the electromagnetic interaction into the free-particle system. As we shall see, the gauge principle will provide us with a deeper understanding of this coupling scheme: To overcome the fact that Eq. (5) is not invariant under the gauge transformation (4), we augment (4) by a space-time-dependent phase transformation of the wavefunction

$$\psi(\mathbf{x},t) \to e^{-ie\theta(\mathbf{x},t)}\psi(\mathbf{x},t), \qquad (7)$$

so that the combinations $(\partial/\partial t - ie\phi)\psi$ and $(\nabla + ie\mathbf{A})\psi$ will have simple transformation properties (the same as ψ) and Eq. (5) will be invariant. We then turn the argument around and reinterpret gauge invariance [which is now the symmetry under the combined transformations of Eqs. (4) and (7)] as follows: The Schrödinger equation for a free charged particle has a global U(1) symmetry (i.e., it is invariant under a phase transformation with θ being independent of **x** and t); to have the freedom of choosing phases of wavefunction locally (i.e., to have a local symmetry) one must introduce some kind of force field, the gauge field, as in Eq. (6). Namely, because of the derivative operators acting on the gauge function, the free Schrödinger equation is no longer invariant under the U(1) transformation with a space-time-dependent phase $\theta(x)$. The gauge potentials ϕ and **A** with the required gauge transformation property (4) are needed so that the extra $\partial\theta/\partial t$ and $\nabla\theta$ factors will cancel. In this way, local phase symmetry completely dictates the electromagnetic coupling of a charged particle as in Eq. (6); and we can look upon the introduction of electromagnetic interaction as resulting from "*gauging*" the U(1) symmetry, i.e., changing the global U(1) symmetry to a local one. This possibility of introducing dynamics by requiring certain symmetry to hold locally is commonly referred to as the *gauge principle*. (The gauging of a symmetry can be motivated in turn by the argument that fundamental global symmetry is unnatural from the viewpoint of relativity since global symmetry would imply that space-like separated parts of a system are not completely independent.)

Why do we call this a gauge symmetry? The term *Eichinvarianz* (gauge invariance), meaning the invariance under a change of the scale (the gauge), was coined by Weyl in 1919 in the framework of his attempt to geometrize the electromagnetic interaction and to construct in this way a unified geometrical theory of gravity and electromagnetism. (Recall Einstein's geometric theory of gravity, general relativity, was published in 1916.) A description in English of his original proposal [Ann. Phys. **59**, 101 (1919)] may be found in

1. **Space-Time-Matter**, H. Weyl, translated by H. L. Brose (Dover, New York, 1951), Chap. IV, Sec. 35, p. 282. The original German edition was published in 1921. (I)

All this was, of course, before the emergence of modern quantum mechanics in 1925. Here, a key concept was to identify the dynamical variables of energy and momentum with the operators $i(\partial/\partial t)$ and $-i\nabla$, and, in the presence of electromagnetic field, with $[i(\partial/\partial t) + e\phi]$ and $(-i\nabla + e\mathbf{A})$ as in the minimal substitution rule (6). In this context, V. Fock [Z. Phys. **39**, 226 (1926)] discovered that the relativistic scalar wave equation (the "Klein–Gordon equation") was invariant under the combined transformation (4) and (7); he called it the "gradient transformation." F. London [Z. Phys. **42**, 375 (1927)] observed that, if the i is dropped from the phase factor in (7), this phase transformation becomes a scale transformation and Eqs. (4) and (7) can then be related to Weyl's old Eichtransformation. However, when Weyl finally worked out this approach [Z. Phys. **56**, 330 (1929)] he retained his original terminology of gauge invariance:

2. **The Theory of Groups and Quantum Mechanics**, H. Weyl, translated by H. P. Robertson (Dover, New York, 1950), Chap. II, Sec. 12, p. 100; Chap. IV, Sec. 5, p. 210. The original German edition (revised) was published in 1930. (I)

Subsequently, phase symmetry and gauge symmetry were often used interchangeably: A global phase transformation was called the *gauge transformation of the first kind*, and a local phase transformation was called the *gauge transformation of the second kind*. In more recent years (the past two decades), the conventional usage in theoretical physics has been such that gauge transformation means gauge transformation of the second kind, and *local symmetry* and *gauge symmetry* are synonymous.

Brief accounts of the early history of the development of the gauge-invariance concept are included in the following articles.

3. **General Principle of Quantum Mechanics**, W. Pauli, translated by P. Achuthan and K. Venkatesan (Springer-Verlag, Berlin, 1980), Secs. 4 and 21. The original German edition was published in *Handbuch der Physik* **24**, edited by Geiger and Scheel (1933). (I)
4. **"Gauge Fields,"** C. N. Yang, in *Proceedings of the 6th Hawaii Topical Conf. Part. Phys.* edited by P. N. Dobson *et al.* (University Press, Honolulu, 1975). (I)

That the electromagnetic potentials represent the more fundamental dynamical variables in quantum mechanics is most clearly demonstrated in the *Aharonov–Bohm effect.*

5. **"Significance of Electromagnetic Potentials in the Quantum Theory,"** Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485–491 (1959). (A)
6. **"Further Considerations on Electromagnetic Potentials in the Quantum Theory,"** Y. Aharonov and D. Bohm, Phys. Rev. **123**, 1511–1524 (1961). (A)

This famous case of physics being affected by potentials has been discussed in numerous instances. Notable examples include:

7. **The Feynman Lectures in Physics**, R. P. Feynman, R. B. Leighton, and M. Sands (Addison-Wesley, Reading, MA, 1964), Vol. II, Sec. 15-5. (E)
8. **"The Aharonov–Bohm Effect: Why it Cannot be Eliminated from Quantum Mechanics,"** M. Peshkin, Phys. Rep. **80**, 375–386 (1981). (I)

9. **Modern Quantum Mechanics,** J. J. Sakurai (Addison-Wesley, Reading, MA, 1985). (I)

The simple U(1) phase transformations of Eq. (7) are commutative, and the corresponding symmetry group is said to be an Abelian group. Modern gauge theory begins with the classic paper by Yang and Mills who first realized that one can extend the above-discussed strategy for constructing a theory with local Abelian symmetry (quantum electrodynamics) to the case of a non-Abelian symmetry group (i.e., a group with noncommutative multiplication law). The case of SU(2) isospin symmetry was explicitly worked out. The structure of such a theory seems to be much richer than the Abelian gauge theory. These new features have led to many interesting developments.

*10. **"Conservation of Isotopic Spin and Isotopic Gauge Invariance,"** C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191–195 (1954). (A)

An early attempt at non-Abelian gauge theory was made by Klein in the context of unified theories of gravity and electromagnetism in a five-dimensional space-time:

11. **"On the Theory of Charged Fields,"** O. Klein, in *New Theories in Physics*, proceedings of a symposium held in Warsaw, 1938. It is reprinted in Okun (Ref. 22), 269–285. (A)

For an account of other works on non-Abelian gauge theories, independent of the Yang–Mills discovery, but unpublished, see, for examples:

12. **Inward Bound: Of Matter and Forces in the Physical World,** A. Pais (Oxford U. P., New York, 1986), Sec. 15(e), pp. 341–346, and Sec. 21(e), pp. 580–620. (E)

It is within this general framework of the Yang–Mills theory that elementary particle theory has made great progress in the past two decades. We now have, for the first time, a comprehensive theory of particle interactions: The strong interaction is described by the "color" gauge group of SU(3); the unified theory of electromagnetic and weak interactions is based on the weak gauge group of $SU(2) \times U(1)$. This *standard model* is at present well supported by all the experimental tests. Furthermore, realistic and credible attempts have been made at *grand unification* in which the above-mentioned two parts of the standard model are contained in a larger gauge group [such as SU(5)] with one gauge coupling.

Gauge invariance, the standard model, and grand unification theories are discussed in an elementary way in

*13. **"Unified Theories of Elementary Particle Interactions,"** S. Weinberg, Sci. Am. **231** (7), 50–59 (1974). (E)

*14. **"Gauge Theories of the Forces between Elementary Particles,"** G.'t Hooft, Sci. Am. **242** (6), 104–138 (1980). (E)

*15. **"Unified Theory of Elementary Particle Forces,"** H. Georgi and S. L. Glashow, Phys. Today **33** (9), 30–39 (1980). (E)

16. **Particle Physics—The Quest for the Substance of Substance,** L. B. Okun (Harwood, London, 1985). (E)

The 1979 Nobel lectures contain excellent intermediate level summary reviews (as well as perspectives for the future):

17. **"Conceptual Foundation of the Unified Theory of Weak and Electromagnetic Interactions,"** S. Weinberg, Rev. Mod. Phys. **52**, 515–524 (1980). (I)

18. **"Gauge Unification of Fundamental Forces,"** A. Salam, Rev. Mod. Phys. **52**, 525–538 (1980). (I)

19. **"Towards a Unified Theory: Thread in a Tapestry,"** S. L. Glashow, Rev. Mod. Phys. **52**, 539–543 (1980). (I)

For pedagogical reviews at the intermediate level we have

20. **An Elementary Primer for Gauge Theory,** K. Moriyasu (World Scientific, Singapore, 1983). (I)

21. **Gauge Theories in Particle Physics,** I. J. R. Aitchison and A. J. G. Hey (Hilger, Bristol, 1982). (I)

22. **"Introduction to Gauge Theories,"** L. B. Okun, Surv. High Energ. Phys. **5**, 199–285 (1986). (I)

Gottfried and Weisskopf have written a set of books that emphasize modes of intuitive thought and semiquantitative calculations; hence these books, although deep, may also be classified as "intermediate" level references:

23. **Concepts for Particle Physics, I & II,** K. Gottfried and V. F. Weisskopf (Clarendon, Oxford, 1986). (I)

At the intermediate level, we have also the two recently published undergraduate particle physics textbooks:

24. **Introduction to Elementary Particles,** D. Griffiths (Harper & Row, New York, 1987). (I)

24a. **Modern Elementary Particle Physics,** G. Kane (Addison-Wesley, Reading, MA, 1987). (I)

At a more advanced level, but without detailed exposition on quantum field theory, we have:

25. **Gauge Theories of the Strong, Weak and Electromagnetic Interactions,** C. Quigg (Benjamin/Cummings, Reading, MA, 1983). (A)

26. **Quarks and Leptons: An Introductory Course in Modern Particle Physics,** F. Halzen and A. D. Martin (Wiley, New York, 1984). (A)

For books on quantum gauge field theory with applications in particle physics, we have

27. **Gauge Theory of Weak Interactions,** J. C. Taylor (Cambridge U.P., London, 1976). (A)

28. **Particle Physics and Introduction to Field Theory,** T. D. Lee (Harwood, New York, 1981). (A)

29. **Leptons and Quarks,** L. B. Okun (North-Holland, Amsterdam, 1982). (A)

30. **Quarks, Leptons and Gauge Fields,** K. Huang (World Scientific, Singapore, 1982). (A)

31. **Gauge Theory of Elementary Particle Physics,** T. P. Cheng and L. F. Li (Clarendon, Oxford, 1984). (A)

32. **Introduction to Gauge Field Theories,** M. Chaichian and N. F. Nelipa (Springer-Verlag, Berlin, 1984). (A)

33. **Gauge Field Theories,** P. Frampton (Benjamin/Cummings, Reading, MA, 1987). (A)

Along with these standard textbooks and monographs, we have the classic reviews:

34. **"Gauge Theories,"** E. Abers and B. W. Lee, Phys. Rep. **9C**, 1–141 (1973). (A)

35. **"Hidden Gauge Symmetry,"** L. O'Raifeartaigh, Rep. Prog. Phys. **42**, 159–223 (1979). (A)

and the set of very illuminating and helpful summer school lectures by Coleman:

36. **Aspects of Symmetry, Selected Erice Lectures,** S. Coleman (Cambridge U.P., London, 1985). (A)

Just about all the above-listed books contain extensive references on gauge theories. (The bibliography in Ref. 25 is particularly detailed and useful.) We urge that readers consult these lists, together with this Resource Letter, in their approach to this immense area of study. In this connection, we would also like to draw the reader's attention to another recent summary review covering similar grounds as this one:

37. **"An Introduction to Standard Model Physics,"** J. L. Rosner, in *TASI 87*, Proc. Theor. Adv. Study Inst.—Santa Fe (World Scientific, Singapore, 1988). (I)

Another helpful reference to note is the following one:

38. **Constructing Quarks—A Sociological History of Particle Physics,** A. Pickering (University of Chicago Press, Chicago, 1984) (E),

which contains a lucid (nontechnical) description of the standard model and its detailed bibliography also lists titles of the cited papers.

Reprint Book: Articles marked with an asterisk (*) will be included in an accompanying Reprint Book. Keeping the Reprint Book of such an extensive field like gauge theories to a reasonable size means that the selection must necessarily be very incomplete. Since one of our criteria is the accessibility of the articles to nonspecialists, the resulting selection tends to emphasize the gauge theory applications in elementary particle physics. In fact, we reprint none of the original papers on the more theoretical topics of spontaneous symmetry breaking, Higgs mechanism, quantization and renormalization of Yang–Mills theory, etc. In this connection, we recommend the lectures on these topics by Coleman (which are unfortunately a bit too long to be included in the reprint volume):

39. **"Renormalization and Symmetry: A Review for Non-Specialists,"** S. Coleman, in Ref. 36, 99–112. (A)

40. **"Secret Symmetry: An Introduction to Spontaneous Symmetry Breakdown and Gauge Fields,"** S. Coleman, in Ref. 36, pp. 113–184. (A)

Even for those topics included in the Reprint Book there is no serious attempt to have a "balanced" coverage. Once again, what we deem to be the "readability" of an article (not to mention our own familiarity and taste) is an important factor in the selection. Fortunately, such a defect can be readily rectified, since the reader can consult the more complete anthology already in existence:

41. **Selected Papers on Gauge Theory of Weak and Electromagnetic Interactions**, edited by C. H. Lai (World Scientific, Singapore, 1982). (A)

Classic papers on grand unification can be found in Ref. 183. And the first papers developing the gauge-invariance concept by Weyl, Fock, and London are reprinted (in German) in a recent publication (Okun, Ref. 22), that the reader may find easier to locate.

Mini Study Guide: As the area covered in this Resource Letter is extensive and many of the cited references are necessarily at an advanced level, a suggested short study guide may be of use to some nonspecialists. Generally speaking, the quickest way to launch a systematic study program is to start with one of the gauge theory textbooks. For a reader with a preparation of undergraduate level physics, we recommend Aitchison and Hey (Ref. 21), supplemented with Griffiths (Ref. 24) and/or Kane (Ref. 24a), especially for the necessary background material in particle physics. To get a deeper understanding, one should then read the important articles collected in the Reprint Book, and study Gottfried and Weisskopf (Ref. 23) for, among other things, its discussion of QCD, Coleman (Ref. 40) for spontaneous symmetry breaking, Quigg (Ref. 25) for electroweak theory. To move on to the quantum gauge field theory level, we suggest, modesty permitting, our own Ref. 31 as a starter.

In the rest of this Resource Letter, we attempt to provide a collection of more specific references useful to nonspecialists wishing to study gauge theories and their applications to elementary particle physics. In Sec. II we shall concentrate on those general properties of this whole class of quantum field theories. In Sec. III gauge theory applications, from QED and the standard model to grand unification, will be discussed.

II. GAUGE THEORIES: GENERAL PROPERTIES

A. Abelian gauge symmetry

We have already discussed electromagnetism as an Abelian gauge interaction. Here, we shall merely restate it in the more convenient relativistic notation of four-vectors:

$$(t,\mathbf{x}) = x^\mu, \quad (\phi,\mathbf{A}) = A^\mu, \quad (\rho,\mathbf{j}) = j^\mu$$

and the tensor:

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix}.$$

The relation between electromagnetic potential A^μ and fields $F^{\mu\nu}$ of Eq. (3) can be compactly expressed as

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu,$$

which is manifestly invariant under the gauge transformation (4):

$$A^\mu \to A^\mu + \partial^\mu\theta. \tag{8}$$

We will use the Lagrangian formalism. This way the symmetry of the system is most easily implemented. Thus, for a free electron, we have (summing over repeated indices $\mu,\nu = 0,...,3$)

$$\mathscr{L} = \bar\psi(i\gamma^\mu\partial_\mu - m)\psi, \tag{10}$$

which leads to the Dirac equation, and for the electromagnetic fields,

$$\mathscr{L} = -\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} + A^\mu j_\mu, \tag{11}$$

which leads to the Maxwell equations through the standard variational principle.

From Eq. (10) it is clear that a free electron has a global U(1) symmetry,

$$\psi(x) \to e^{-ie\theta}\psi(x), \tag{12}$$

where θ is some constant. To gauge this symmetry, i.e., taking the phase factor to be a function of space-time as in Eq. (7), we simply replace the ordinary derivative ∂^μ in Eq. (10) by *covariant derivative*:

$$D^\mu = \partial^\mu + ieA^\mu. \tag{13}$$

This, of course, introduces the electromagnetic interaction through the minimal coupling of (6). The significance of the covariant derivative is that while the ordinary derivative transforms as $\partial^\mu\psi \to e^{-ie\theta(x)}[\partial^\mu\psi - ie(\partial^\mu\theta)\psi]$, the combination $D^\mu\psi$ has the same homogeneous transformation as the field ψ itself:

$$[D^\mu\psi(x)] \to e^{-ie\theta(x)}[D^\mu\psi(x)]. \tag{14}$$

Thus the local $U(1)$ symmetry can be maintained in the combination of $\bar\psi D_\mu\psi$. We now want to write the kinetic energy term for A^μ, which should involve the derivative $\partial^\mu A^\nu$. That the simplest gauge-invariant term involving the $\partial^\mu A^\nu$ factor is $F^{\mu\nu}$ of Eq. (8) suggests that the full QED Lagrangian density is

$$\mathscr{L} = -\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar\psi(i\gamma^\mu D_\mu - m)\psi. \tag{15}$$

The $(-\tfrac{1}{4})$ factor is needed to reproduce the Lagrangian for the electromagnetic fields in the standard form Eq. (11). This simple Lagrangian for the case of $\psi(x)$ corresponding to the electron field actually works remarkably well; (see discussion in Sec. III A).

Note that gauge invariance by itself will allow interaction terms such as $(F^{\mu\nu}F_{\mu\nu})^n$ with $n > 1$ and $\bar\psi\sigma_{\mu\nu}\psi F^{\mu\nu}$ with $\sigma_{\mu\nu} = (i/2)[\gamma^\mu,\gamma^\nu]$, etc. But we can exclude all such terms by requiring the theory to be "*renormalizable*." Roughly speaking, this means that we keep in the Lagrangian density only terms having (mass) dimension less than or equal to four. [The spin-0 and spin-1 fields have the same dimension as the derivative operator and the spin-1/2 field has dimension 3/2, etc. Thus $(FF)^n$ terms have dimension $4n$ and $\bar\psi\sigma\psi F$ dimension 5. Since the Lagrangian density has dimension 4, such nonrenormalizable terms must have negative dimensional (coupling) coefficients that will bring about uncontrollable divergences in the loop diagrams.] Thus the requirement of gauge invariance and renormalizability completely fix the form of QED. This way of viewing QED has been advocated by Schwinger:

42. **"The Theory of Quantized Fields II,"** J. Schwinger, Phys. Rev. **91**, 713–728 (1953). (A)

The introduction of a "compensating" gauge field via the covariant derivative (13) has a natural geometric interpretation: If the phase convention of the electron wavefunction is chosen independently at each space-time point, then the gauge potential factor in Eq. (13) simply represents (i.e., parametrizes) the "extra" phase change, due to the change of phase convention, between two neighboring points separated by dx^μ. Thus, when an electron travels a finite distance, say from point 1 to 2 via a path C, it undergoes a phase change of

$$\Phi_{12}^{(C)} = \exp\left(ie \int_C dx^\mu A_\mu(x)\right). \tag{16a}$$

From this expression, we see that a gauge transformation in Eq. (8) represents merely a change of the (relative) phase convention at points 1 and 2. This phase (16a) is not measurable; only its difference when compared with another phase factor resulting from a different path, say C', is physical (as in a double-slit experiment). This physical phase can be expressed as a closed-path integral:

$$\Delta\Phi = \Phi_{12}^{(C)} - \Phi_{12}^{(C')} = \exp\left(ie \oint dx^\mu A_\mu(x)\right), \tag{16b}$$

which is unchanged under an arbitrary gauge transformation. (In this connection, we note that, using the Stokes theorem, the right-hand side can be written as a surface integral over the field tensor $F_{\mu\nu}$, which is manifestly gauge invariant.) Thus, in this approach, the form of gauge transformation (8) comes out very naturally. This *nonintegral* (i.e., *path-dependent*) *phase* viewpoint of gauge symmetry has been advocated in several publications, one being Ref. 43 (this one actually on non-Abelian symmetry):

43. **"Integral Formalism for Gauge Fields,"** C. N. Yang, Phys. Rev. Lett. **33**, 445–449 (1974). (A)

This, for example, is also the approach taken in the nontechnical presentation (Ref. 14) by 't Hooft.

B. Non-Abelian gauge symmetry

Modern gauge theory was started in 1954 by Yang and Mills, who constructed the SU(2) gauge theory (Ref. 10). The extension to theories based on general non-Abelian gauge symmetry groups has been studied by Gell-Mann and Glashow:

44. **"Gauge Theories and Vector Particles,"** S. L. Glashow and M. Gell-Mann, Ann. Phys. **15**, 437–467 (1967). (A)

For the non-Abelian theory, the single gauge function $\theta(x)$ in the phase transformation (7) corresponding to the Abelian U(1) symmetry is replaced by a set of gauge functions $\theta_a(x)$, $a = 1,2,...,N$. The transformation is extended to a matrix multiplication

$$\psi_i(x) \to \exp[i\theta_a(x)T_a]_{ij}\psi_j \equiv [U(x)]_{ij}\psi_j(x), \tag{17}$$

where T_a's are matrices with component indices i and j ranging over dimension of the ψ_i multiplet. They satisfy the commutation relation of the symmetry group

$$[T_a, T_b] = if_{abc}T_c,$$

where f_{abc} are the structure constants of the group.

To construct the covariant derivative, one introduces a set of gauge fields $A_a^\mu(x)$, $a = 1,...,N$, through the minimal coupling

$$[D^\mu\psi(x)]_i = \partial^\mu\psi_i(x) - ig[T_aA_a^\mu(x)]_{ij}\psi_j(x), \tag{18}$$

where g is the gauge coupling just as e is the coupling for the electromagnetic $U(1)$ gauge theory. In order for the covariant derivative $(D^\mu\psi)_i$ to have the same transformation as ψ_i itself,

$$[D^\mu\psi(x)] \to U(x)[D^\mu\psi(x)],$$

the gauge fields $A_a^\mu(x)$ are required to transform as follows:

$$[T_aA_a^\mu(x)] \to U(x)[T_aA_a^\mu(x)]U^{-1}(x) - (i/g)[\partial^\mu U]U^{-1}(x). \tag{19}$$

The form of the first term on the right-hand side [i.e., the surviving term when $U(x)$ is a constant matrix] indicates that, in contrast to the Abelian theory, the gauge fields A_a^μ transform nontrivially under the symmetry group. This is because the matrices T_a's do not commute. The generalization of the field strength tensor of Eq. (8) turns out to be

$$F_a^{\mu\nu} = \partial^\mu A_a^\nu - \partial^\nu A_a^\mu - gf_{abc}A_b^\mu A_c^\nu, \tag{20}$$

which transforms simply as

$$[T_aF_a^{\mu\nu}(x)] \to U(x)[T_aF_a^{\mu\nu}(x)]U^{-1}(x).$$

The presence of the quadratic $(A)^2$ term in $F_a^{\mu\nu}$ is again due to the non-Abelian nature of the symmetry group. Now we can write down the gauge-invariant Lagrangian density as

$$\mathscr{L} = -\tfrac{1}{4}F_a^{\mu\nu}F_{a\mu\nu} - \bar{\psi}(i\gamma^\mu D_\mu - m)\psi. \tag{21}$$

While similar in appearance to the Abelian QED Lagrangian of Eq. (15), the non-Abelian equation (21) contains cubic and quartic terms in the gauge fields A_a^μ. Thus the corresponding field equations are nonlinear. In more physical terms, while the photon field is electrically neutral and interacts only with charged "matter-particles" such as the electron, the non-Abelian gauge fields carry the gauge symmetry charges themselves and have self-interactions, which are also determined by the gauge symmetry. Here, we see even more clearly the remarkable property of gauge symmetry that it not only relates masses and coupling constants, as the usual global symmetry, but also determines the gauge field interactions with other particles as well as their self-interactions.

In this respect, we also see the similarity of the Yang–Mills theory to the general theory of relativity. In the latter case, the (nonlinear) dynamics of gravitational interactions are to a large extent fixed by the requirement of invariance with respect to the general coordinate transformation. In fact, general relativity may be viewed as a gauge theory. Instead of the geometric formulation of Einstein within the basic mathematical framework of Riemannian curved space-time, one can adopt a more algebraic approach in which the theory arises from gauging the Lorentz symmetry group. This formulation was presented first by Utiyama:

45. **"Invariant Theoretical Interpretation of Interaction,"** R. Utiyama, Phys. Rev. **101**, 1597–1607 (1956) (A),

which is extended to the inhomogeneous Poincaré group by Kibble:

46. **"Lorentz Invariance and the Gravitational Field,"** T. W. B. Kibble, J. Math. Phys. **2**, 212–221 (1961). (A)

In this approach, gravitation is introduced via a gauge potential Γ^μ, which corresponds to $g(T_a A^\mu_a)$ in Eq. (18) with the matrix indices i and j being themselves the four-dimensional space-time indices: $(\Gamma^\mu)^\beta_\alpha = \Gamma^{\mu\beta}_\alpha$. Thus the covariant derivative of a vector $V_\alpha(x)$ is given by

$$D^\mu V_\alpha(x) = \partial^\mu V_\alpha(x) + \Gamma^{\mu\beta}_\alpha(x) V_\beta(x).$$

The gauge potential Γ_μ is called the *affine connection*. One can easily show that the field strength tensor corresponding to $g(T_a F^{\mu\nu}_a)^\beta_\alpha$ is just the Riemann curvature tensor $R^{\mu\nu\beta}_\alpha$. A word of caution: Gravitation has more structure than the gauge structure as indicated here. For example, it has a metric substructure that may be viewed as a sort of prepotential from which Γ_μ may be derived (i.e., as the *Christoffel symbol*) if a "*no-torsion*" assumption is made, etc. Our purpose here is merely to show the universality of the gauge principle. For a reader interested in the various deep issues related to the gauge theoretical approach to gravity, he or she may start tracing the relevant literature from a recent review:

47. "**The Gauge Treatment of Gravity**," D. Ivanenko and G. Sardanashvily, Phys. Rep. **94**, 1–45 (1983). (A)

In the following, we shall restrict our discussion to compact gauge groups and their applications in elementary particle physics.

One property of the non-Abelian Lagrangian (21) is the same as the Abelian case of (15): It also lacks the mass term $(A)^2$. Thus it seems that the quanta of the gauge fields are always massless. This zero-mass property of the gauge quanta (which implies a long-range interaction) stands as one of the hurdles in any application of the Yang–Mills theory. Four interactions are known in nature: Gravity and electromagnetism are long ranged and have zero-mass quanta (graviton and photon); they are indeed described by gauge theories—in fact, they inspired the formulation of the gauge-invariance concept in the first place. The remaining two interactions, strong and weak, are short ranged and it is not clear at all how the Yang–Mills theory can be applied to such interactions.

The resolution of this fundamental problem came about after years of theoretical developments. In fact, there are two distinctive explanations for the short-range nature of these two interactions. Relevant to the weak interaction, it has been demonstrated that when the gauge symmetry is "hidden" because of *spontaneous symmetry breaking*, a gauge quantum acquires a mass. Relevant to the strong interaction, it was eventually learned that non-Abelian gauge theory, unique among all other quantum field theories, is *asymptotically free*, i.e., the effective coupling vanishes asymptotically at short distances. This offers the possibility that the coupling grows at long distance. The observed strongly interacting baryons and mesons are actually not elementary and the observed low-energy strong interactions are not fundamental but are remnants of the fundamental "color" gauge interactions among the more elementary hadronic constituents of quarks and gluons (much in the same way that the molecular and chemical forces are the remnants of the more fundamental electromagnetic interactions among atomic constituents). The color gauge interactions become strong at long distances and they are confined to short distances ("*infrared slavery*"). In Secs. II C and II D we shall discuss these two properties of spontaneous symmetry breaking and asymptotic freedom in turn.

C. Spontaneous symmetry breaking

A symmetry can be hidden if the dynamics are such that the ground state (the vacuum) is not itself a symmetric (i.e., a singlet) state. Consequently, even though the basic interaction, as described by the equation of motion, is invariant under the symmetry transformations, the choice of one vacuum state from all the possible degenerate vacua breaks the symmetry, and the invariance is no longer manifest in the degeneracy of the physical states. Such a hidden symmetry situation is said to display *spontaneous symmetry breaking*. One of the leading pioneers of this theoretical development (first in the context of global symmetry) is Nambu:

48. "**Axial Vector Current Conservation in Weak Interactions**," Y. Nambu, Phys. Rev. Lett. **4**, 380–382 (1960). (A)
49. "**Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I**," Y. Nambu and G. Jona-Lasino, Phys. Rev. **122**, 345–358 (1961). (A)
50. "**Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. II**," Y. Nambu and G. Jona-Lasino, Phys. Rev. **124**, 246–254 (1961). (A)

That spontaneous breaking of continuous symmetry implies the existence of massless, spinless particles (the *Nambu–Goldstone bosons*) was further studied, and proven with various degrees of rigor and generality, in

51. "**Field Theories with 'Superconductor' Solutions**," J. Goldstone, Nuovo Cimento **18**, 154–164 (1961). (A)
52. "**Broken Symmetries**," J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965–970 (1961). (A)

For a useful review of spontaneous symmetry breaking and conservation laws, see

53. "**Broken Symmetries and the Goldstone Theorem**," G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, in *Advances in Particle Physics*, edited by R. Cool and R. Marshak, Vol. 2 (Interscience, New York, 1968), pp. 567–708. (A)

One of the most important applications of spontaneously broken global symmetry is the *chiral symmetry* of strong interactions in which the pseudoscalar mesons, π's, K's, and η behave approximately like Goldstone bosons. The study of chiral symmetry in strong interaction is commonly referred to as *current algebra*. Useful references on this subject may be found in

54. **Current Algebra and Application to Particle Physics**, S. L. Adler and R. F. Dashen (Benjamin, New York, 1968). (A)
55. **Currents in Hadron Physics**, Y. de Alfaro, S. Fubini, G. Furlan, and C. Rosetti (North-Holland, Amsterdam, 1973). (A)
56. **Current Algebra and Anomalies**, S. B. Trieman, R. Jackiw, B. Zumino, and E. Witten (Princeton U. P., Princeton, NJ, 1986). (A)

When spontaneous symmetry breaking takes place in theories with local symmetries, something rather dramatic happens: The zero-mass Goldstone bosons combine with the zero-mass vector gauge bosons to form massive vector particles. Namely, in a situation of spontaneous broken local symmetry, the gauge boson gets its mass from the interaction of gauge bosons with the spin-zero bosons. Such a scenario (especially when it involves elementary scalar fields) has come to be called the *Higgs mechanism* and the scalars, the *Higgs particles*.

Anderson first pointed out that several cases in nonrelativistic condensed matter physics may be interpreted as due to "massive photons":

57. "**Random-Phase Approximation in the Theory of Superconductivity**," P. W. Anderson, Phys. Rev. **112**, 1900–1916 (1958). (A)

After the above-mentioned development in the study of spontaneous broken symmetry, he returned to this gauge symmetry case:

58. "**Plasmons, Gauge Invariance, and Mass**," P. W. Anderson, Phys.

Rev. **130**, 439–442 (1963). (A)

That dynamical effects can evade the local symmetry predictions of a zero-mass gauge boson was suggested earlier in

59. "**Gauge Invariance and Mass**," J. Schwinger, Phys. Rev. **125**, 397–398 (1962). (A)

Schwinger also worked out an explicit example of such a situation in the case of two-dimensional QED with massless fermions

60. "**Gauge Invariance and Mass. II**," J. Schwinger, Phys. Rev. **128**, 2425–2429 (1962). (A)

Generation of gauge boson masses via spontaneous symmetry breaking in four-dimensional relativistic field theories was worked out nearly simultaneously in the following papers:

61. "**Broken Symmetry and the Mass of Gauge Vector Mesons**," F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321–323 (1964). (A)
62. "**Broken Symmetries and the Masses of Gauge Bosons**," P. W. Higgs, Phys. Rev. Lett. **13**, 508–509 (1964). (A)
63. "**Global Conservation Laws and Massless Particles**," G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13**, 585–587 (1964). (A)

Further elaboration and general group-theoretical analysis were offered in

64. "**Spontaneous Symmetry Breakdown with Massless Bosons**," P. W. Higgs, Phys. Rev. **145**, 1156–1163 (1966). (A)
65. "**Symmetry Breaking in Non-Abelian Gauge Theories**," T. W. B. Kibble, Phys. Rev. **155**, 1554–1561 (1967). (A)
66. "**General Theory of Broken Local Symmetries**," S. Weinberg, Phys. Rev. D **7**, 1068–1082 (1973). (A)
67. "**Group Theory of the Spontaneous Broken Gauge Symmetries**," L. F. Li, Phys. Rev. D **9**, 1723–1739 (1974). (A)

The Higgs mechanism of breaking a gauge symmetry is discussed in an elementary way (via a geometric approach) by 't Hooft (Ref. 14). Also see:

68. "**Breaking of Gauge Symmetry: A Geometric View**," K. Moriyasu, Am. J. Phys. **48**, 200–204 (1980). (I)

For a more detailed introduction, the reader is referred to Coleman (Ref. 40) and to:

69. "**Spontaneous Symmetry Breaking, Gauge Theories, the Higgs Mechanism, and All That**," J. Bernstein, Rev. Mod. Phys. **46**, 7–48 (1974). (I)
70. "**Spontaneous Symmetry Breaking**," P. W. Higgs, in *Phenomenology of Particles at High Energies*, 14th Scottish Univ. Sum. Sch. Phys. 1973, edited by R. L. Crawford and R. Jennings (Academic, London, 1974), 529. (A)

Many of the gauge theory reviews listed in the Introduction contain discussions of Higgs mechanism; see, for example: Aitchison and Hey (Ref. 21), Quigg (Ref. 25), Abers and Lee (Ref. 34), Cheng and Li (Ref. 31), Frampton (Ref. 33), etc.

D. Asymptotic freedom

In a quantum field theory, to obtain physical results a program of *renormalization* must be carried out. This prescription consistently isolates and removes all the infinities encountered in the perturbation calculations from the physically measurable quantities. It has been of utmost importance to the development of relativistic quantum field theory. The difficult program to implement this in QED was carried out in the late 1940s (see further discussion in Sec. II E), and has led to predictions that are in spectacular agreement with experimental measurements (see discussion in Sec. III A).

We can illustrate the physical basis for renormalization with the simple example of a charged particle moving inside a solid. It is not difficult to appreciate the fact that due to the interaction of the charged particle with the lattice the effective charge and mass (e,m) that determine its response to an externally applied electric field are different from the charge and mass measured outside the solid. In other words, the particle's charge and mass are changed (renormalized) from (e_0,m_0) to (e,m). In this simple case, one can, in principle, measure both (e,m) and (e_0,m_0) by switching on and off the interaction (i.e., by placing the particle inside or outside the solid). Clearly their differences are finite since both sets of quantities are measurable and finite. For the relativistic field theory, the situation is the same except for two important distinctions. First, renormalization due to interaction is generally infinite. Second, there is no way to switch off the interaction; hence quantities in the absence of interaction, called the "unrenormalized" or the "bare" quantities (e_0,m_0), are not measurable. Thus we can allow the bare quantities to be infinite in such a way that the physical measurable quantities are finite. In QED, charge renormalization is just "vacuum polarization," since the ground state (the vacuum) in QED should be thought of as a medium containing virtual electron–positron pairs. The renormalized charge is given by e_0/ϵ, with ϵ being the dielectric constant. Since ϵ depends on frequency (or energy, or distance), we can introduce the notion of an effective coupling $e(r) \equiv e_0/\epsilon(r)$ that governs the strength of electrodynamic interactions at a distance r. The theoretical tool to make such studies in relativistic quantum field theory is the *renormalization group*, as first done in:

71. "**Quantum Electrodynamics at Small Distances**," M. Gell-Mann and F. E. Low, Phys. Rev. **95**, 1300–1312 (1954). (A)

Renormalization groups were discovered by E. C. G. Stueckelberg and A. Petermann [Helv. Phys. Acta **26**, 499 (1953), in French]. Their role in the Gell-Mann–Low analysis was discussed and its general formulation studied by Bogoliubov and Shirkov [Doklady. Akad. Nauk. **103**, 203 (1955), JETP **30**, 77 (1955)]. An exposition of this formulation can be found in:

72. **An Introduction to Quantized Fields**, N. N. Bogoliubov and D. V. Shirkov (Wiley–Interscience, New York, 1959), 1st ed. (A)

The resurgence of interest in the application of the renormalization group to study the effect of a scale change in a theory is largely brought about by the work of Wilson:

73. "**Renormalization Group and Strong Interactions**," K. G. Wilson, Phys. Rev. D **3**, 1818–1846 (1971). (A)

For a nontechnical account, see

74. "**Problems in Physics with Many Scales of Length**," K. G. Wilson, Sci. Am. **241** (8), 158–179 (1979). (E)

For a technical, but very readable, introduction to renormalization groups, see:

75. "**Dilations**," S. Coleman in Ref. 36, 67–98. (A)

In quantum electrodynamics it is easy to understand that the effective charge $e(r)$ decreases as r increases because the production of virtual electron–positron pairs screen the source. However, the corresponding situation involving non-Abelian gauge fields is different. (Quantization and renormalization of Yang–Mills theories will be discussed in the next subsection.) The virtual gauge particle contribution is such that, as r decreases, the effective interaction strength becomes smaller—the theory asymptotically approaches a free field theory and is perturbatively calculable in this limit. That the Yang–Mills theory is *asymptotically free* was discovered by Gross and Wilczek, and by Politzer:

*76. "**Ultraviolet Behavior of Non-Abelian Gauge Theories**," D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343–1346 (1973). (A)

*77. **"Reliable Perturbative Results for Strong Interactions?,"** H. D. Politzer, Phys. Rev. Lett. **30**, 1346–1349 (1973). (A)

That there can be no asymptotic freedom in a theory without non-Abelian gauge fields was proven in the following paper:

78. **"The Price of Asymptotic Freedom,"** S. Coleman and D. J. Gross, Phys. Rev. Lett. **31**, 851–854 (1973). (A).

Asymptotic freedom comes about because the virtual Yang–Mills spin-1 bosons, unlike the Abelian photons, carry gauge symmetry charges themselves and they, unlike the spin-1/2 and spin-0 virtual particles, "antiscreen" the source. An illuminating study of this phenomenon as a paramagnetism of the Yang–Mills system was carried out by Nielson:

79. **"Asymptotic Freedom as a Spin Effect,"** N. K. Nielson, Am. J. Phys. **49**, 1171–1178 (1981). (I)

80. **"The Physics of Asymptotic Freedom,"** K. Johnson, in *Asymptotic Realms of Physics*, edited by A. Guth, *et al.* (MIT Press, Cambridge, MA, 1983). (I)

For general discussions of asymptotic freedom and its profound implication for a quantum field theory, see the reviews by Gross:

*81. **"Asymptotic Freedom,"** D. J. Gross, Phys. Today **40** (1), 39–44 (1987). (E)

82. **"Applications of Renormalization Group Equations in High Energy Physics,"** D. J. Gross, in *Methods in Field Theory*, Proc. of 1975 Les Houches Summer School, edited by R. Balian and J. Zinn-Justin (North-Holland, Amsterdam, 1976). (A)

To the extent that perturbation results are applicable, asymptotic freedom also means that interaction strength increases with distance. This suggests the possibility that this increase will be so large that it will take infinite energy to separate a pair of particles with nonzero gauge symmetry charge to infinity. This long-distance behavior of the field theory has been dubbed "*infrared slavery*" or simply "*confinement*." This is just the quark trapping mechanism needed to explain the nonobservation of free quarks and it can account for the short-range nature of the strong interaction. This proposition was made by a number of authors; we mention Gross and Wilczek (Ref. 76) and:

*83. **"Non-Abelian Gauge Theories of the Strong Interactions,"** S. Weinberg, Phys. Rev. Lett. **31**, 494–497 (1973). (A)

*84. **"Advantages of the Color Octet Gluon Picture,"** H. Fritzsch, M. Gell-Mann, and H. Leutwyler, Phys. Lett. **47B**, 365–368 (1973). (A)

For a very physical discussion of asymptotic freedom and confinement, see Gottfried and Weisskopf (Ref. 23).

E. Quantization and renormalization

There are many books on the subject of quantum field theory. At a more introductory level we have, for example,

85. **Quantum Fields**, N. N. Bogoliubov and D. V. Shirkov (Benjamin/Cummings, Reading, MA, 1983). (I)

86. **Quantum Field Theory**, F. Mandl and G. Shaw (Wiley–Interscience, Chichester, 1984). (I)

For more advanced discussions, but still restricted to the Abelian QED, we have:

87. **The Theory of Photons and Electrons**, J. M. Jauch and F. Rohrlich (Springer, New York, 1976), 2nd ed. (A)

88. **Relativistic Quantum Fields**, J. D. Bjorken and S. D. Drell (McGraw-Hill, New York, 1965). (A)

89. **Relativistic Quantum Theory**, V. B. Beresteskii, E. M. Lifshitz, and L. P. Pitaevskii (Pergamon, New York, 1971). (A)

90. **Introduction to the Theory of Quantum Fields**, N. N. Bogoluibov and D. V. Shirkov (Wiley–Interscience, New York, 1980), 3rd ed. (A)

For discussions of quantum field theories that also include non-Abelian gauge fields, we have already mentioned Abers and Lee (Ref. 34), Coleman (Ref. 36), Taylor (Ref. 27), Lee (Ref. 28), Huang (Ref. 30), Cheng and Li (Ref. 31), Chaichian and Nelipa (Ref. 32), and Frampton (Ref. 33). For books with more of an emphasis on quantum field theory rather than on specific applications, we have:

91. **Field Theory, A Modern Primer**, P. Ramond (Benjamin/Cummings, Reading, MA, 1981). (A)

92. **Gauge Fields**, N. P. Konopleva and V. N. Popov (Harwood, New York, 1981). (A)

93. **Introduction to Quantum Field Theory**, C. Itzykson and J. Zuber (McGraw-Hill, New York, 1980). (A)

94. **Quantum Field Theory**, L. H. Ryder (Cambridge U.P., London, 1985). (A)

For readers interested in a more detailed discussion of the renormalization of Yang–Mills theory, we recommend:

95. **Renormalization—An Introduction to Renormalization, The Renormalization Group, and the Operator-Product Expansion**, J. C. Collins (Cambridge U.P., London, 1984). (A)

Let us highlight some of the key features in the quantization and renormalization of gauge-invariant field theory. We recall that the usual procedure in quantizing a classical system involves replacing the dynamical variables by operators satisfying commutation relations corresponding to the classical Poisson brackets. In the case of gauge theories, the complication arises because not all components of the gauge fields A_μ with $\mu = 0,1,2,3$ are true dynamical variables because they are related by gauge invariance. We can eliminate the unphysical degrees of freedom with a properly imposed constraint. In QED, this constraint may be in the form of $\nabla \cdot \mathbf{A}(x) = 0$ (radiation gauge), or $A_3 = 0$ (axial gauge), etc. Fixing the gauge in this way one sacrifices manifest Lorentz covariance. Alternatively, one can quantize the theory covariantly by carrying along these unphysical degrees of freedom. But covariant quantization (i.e., covariant gauges) inevitably bring in negative norm states in the Hilbert space. (The theory is said to have an "*indefinite metric*.") For an Abelian gauge theory, one can remove such states by imposing the subsidiary condition of

$$(\partial^\mu A_\mu)|\text{phys}\rangle = 0 \tag{22}$$

to select out the positive norm physical state $|\text{phys}\rangle$ and there is no modification of the usual Feynman rules.

These two simple approaches to quantize QED do not seem to work for the non-Abelian gauge theory because here the equations of motion are nonlinear in the gauge fields. In 1967 Faddeev and Popov provided the crucial step of writing down the correct quantization procedure for the non-Abelian gauge theory by using the path integral formalism. Their result can be cast in the form of introducing further unphysical degrees of freedom into the theory: the *Faddeev–Popov ghost particles*. The covariant-gauge Feynman rule can then be derived from an effective Lagrangian:

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_g + \mathscr{L}_{\text{gf}} + \mathscr{L}_{\text{FP}} \tag{23}$$

where $\mathscr{L}_g$ is the gauge-invariant classical Lagrangian; $\mathscr{L}_{\text{gf}}$ is the gauge fixing term, and $\mathscr{L}_{\text{FP}}$ is the Faddeev–Popov ghost term.

96. **"The Feynman Diagrams for Yang–Mills Fields,"** L. D. Faddeev and V. N. Popov, Phys. Lett **25B**, 29–30 (1967). (A)

Thus, in order to quantize the theory in covariant gauges, one must introduce more unphysical degrees of freedom. Even so, one can prove that by restricting attention to the set of gauge-invariant states, *all* spurious states are removed. (This will be discussed further in connection with BRST symmetry below.)

The path integral formalism is now generally recognized as the most convenient method for quantizing the constraint system such as the Yang–Mills theories. Useful introductions to this approach of quantization can be found in Abers and Lee (Ref. 34), Cheng and Li (Ref. 31), Faddeev and Slavnov (Ref. 90), Ramond (Ref. 91), Itzykson and Zuber (Ref. 93), etc. For specific references on path integral quantization, see:

97. **Quantum Mechanics and Path Integrals**, R. P. Feynman and H. R. Hibbs (McGraw-Hill, New York, 1965). (I)

98. **Techniques and Applications of Path Integration**, L. S. Schulman (Wiley, New York, 1981). (I)

As discussed in Sec. II D, to study the physical consequence of the relativistic quantum field theory, one needs a consistent prescription to remove all the divergences from the physically measurable quantities (the "*renormalization program*"). This usually involves a very complicated analysis of the divergence properties of various Feynman diagrams. If all the infinities can be removed by the redefinition of the basic parameters like masses and coupling constants, we say the theory is renormalizable, and we can then predict all the other measurable quantities in terms of these basic parameters. This difficult program was first carried out in quantum electrodynamics; the basic QED quantization and renormalization papers are collected in the following anthology:

99. **Selected Papers in Quantum Electrodynamics**, edited by J. Schwinger (Dover, New York, 1958). (A)

It turns out that there are very few types of theories that are renormalizable. Thus the requirement of renormalizability places a very severe constraint on the possible form of interaction.

One of the important steps in the renormalization program is the implementation of the regularization procedure that renders the divergent integrals finite so that the theory is well defined for further mathematical manipulations. The renormalizability of a theory with symmetry depends critically on the cancellation of divergences as enforced by the symmetry relations, *the Ward identities*, among Green's functions. (References on Ward identities will be given below.) Thus one must have a regularization procedure that guarantees that the Ward identities are satisfied so that the symmetry of the theory is preserved. For the Abelian gauge theory of QED a simple method that preserves the gauge symmetry is the *Pauli–Villars regularization*, involving covariant cutoff of the propagators:

100. **"On the Invariant Regularization in Relativistic Quantum Theory,"** W. Pauli and F. Villars, Rev. Mod. Phys. **21**, 434–444 (1949). (A)

But generalizing this to the Yang–Mills theory is not straightforward. The invention of an entirely different procedure, the *dimensional regularization*, has made the study of renormalization of the non-Abelian gauge theory more manageable. The basic idea is that by going to lower spacetime dimensions, one can make the theory finite and preserve the symmetry at the same time. This elegant regularization procedure was invented by several groups independently; here, we shall only quote the most influential one:

101. **"Regularization and Renormalization of Gauge Fields,"** G. 't Hooft and M. Veltman, Nucl. Phys. B **44**, 182–213 (1972). (A)

For a review of the subject matter, see:

102. **"Introduction to Technique of Dimensional Regularization,"** G. Leibrandt, Rev. Mod. Phys. **47**, 847–876 (1975). (A)

By using the Faddeev–Popov techniques for quantization and the procedure of dimensional regularization, 't Hooft made a systematic study of the renormalization property of the non-Abelian gauge theory. In 1971, he presented his proofs that the Yang–Mills theory is renormalizable, with or without spontaneous symmetry breaking.

103. **"Renormalization of Massless Yang–Mills Fields,"** G. 't Hooft, Nucl. Phys. B **33**, 173–199 (1971). (A)

104. **"Renormalizable Lagrangian for Massive Yang–Mills Fields,"** G. 't Hooft, Nucl. Phys. B **35**, 167–188 (1971). (A)

The result of the second paper was particularly important. Before this, the general expectation was pessimistic with respect to the renormalizability of any massive Yang–Mills theory and this class of quantum field theories was precisely the one thought to have direct physical relevance. Thus the significance of 't Hooft's achievement was appreciated very quickly by the community of particle theorists. This transformed the whole field of elementary particle theory and brought about the renaissance of quantum field theory in the 1970s. A very useful historical account of the evolution of quantum gauge theory was given by Veltman:

105. **"Gauge Field Theories,"** M. Veltman, in *Proceedings of the 6th International Symposium on Electron and Photon Interactions at High Energies* (Bonn, 1973), edited by H. Rollnik and W. Pfeil (North-Holland, Amsterdam, 1974), 429–447. (I)

To conclude this section, we briefly discuss three topics: Ward identities, BRST symmetry, and anomaly. Gauge invariance manifests itself in a set of algebraic relations among Green's functions called Ward identities. These symmetry relations play a crucial role in the renormalization of the theory:

106. **"An Identity in Quantum Electrodynamics,"** J. C. Ward, Phys. Rev. **78**, 182 (1950). (A)

107. **"On the Generalized Ward Identity,"** Y. Takahashi, Nuovo Cimento **6**, 370–375 (1957). (A)

The generalized Ward identities of the non-Abelian gauge theory, which also include those involving only the Yang–Mills fields and the associated ghosts, were first studied by Slavnov and by Taylor:

108. **"Ward Identities in Gauge Theories,"** A. A. Slavnov, Theor. Math. Phys. **10**, 99–104 (1972). (A)

109. **"Ward Identities and Charge Renormalization of the Yang–Mills Field,"** J. C. Taylor, Nucl. Phys. B **33**, 436–444 (1971). (A)

These *Slavnov–Taylor identities*, which contain the full non-Abelian gauge-invariance content of the theory, are indispensible tools in showing the unitarity property of a theory containing ghost loops. Their derivation is very much simplified by making use of the symmetry discovered by Becchi, Rouet, and Stora [Phys. Lett. **52B**, 344 (1974)] and by I. V. Tyupin [Lebedev preprint FIAN No. 39 (1975), in Russian, unpublished]. For a detailed account, see

110. **"Renormalization of Gauge Theories,"** C. Becchi, A. Rouet, and R. Stora, Ann. Phys. **98**, 287–321 (1976). (A)

The effective Lagrangian (23) of the quantum theory is no longer invariant under gauge transformations with arbitrary gauge functions $\theta_a(x)$ because we have fixed the gauge (which, in turn, requires us to add ghost particles). It has, however, a global fermionic symmetry, the BRST symmetry, corresponding to the transformations with $\theta_a(x)$ being proportional to the Faddeev–Popov (fermionic) ghost field. This symmetry is not broken and a physical state must be BRST invariant. In fact, the non-Abelian generalization of the subsidiary condition (22) is simply

$$Q\,|\text{phys}\rangle = 0, \tag{24}$$

where Q is the BRST charge, which generates BRST transformations. For a detailed discussion of this quantization formalism, see:

111. **"Local Covariant Operator Formalism of Non-Abelian Gauge The-**

ories, and Quark Confinement Problem," T. Kugo and I. Ojima, Suppl. Prog. Theor. Phys. **66**, 1–130 (1979). (A)
112. **"A BRST Primer,"** D. Nemeschansky, C. Preitschopf, and M. Weinstein, SLAC-PUB-4422 (1987), to be published. (A)

The "*anomaly*" phenomenon is the situation where the quantum loop corrections break the symmetry of the original (classical) Lagrangian. The most famous case is the *chiral anomaly*. Here, the conservation of axial vector current is modified by the fermion triangle diagram. Realization of an anomaly's significance for quantum field theory began with the current algebra work of Adler, Bell, and Jackiw on the decay $\pi^0 \to \gamma\gamma$:

113. **"Axial Vector Vertex in Spinor Electrodynamics,"** S. L. Adler, Phys. Rev. **177**, 2426–2438 (1969). (A)
114. **"A PCAC Puzzle: $\pi^0 \to \gamma\gamma$ in the σ-Model,"** J. S. Bell and R. Jackiw, Nuovo Cimento **60A**, 47–61 (1969). (A)

The theory of the non-Abelian anomaly was worked out by Bardeen:

115. **"Anomalous Ward Identities in Spinor Field Theories,"** W. A. Bardeen, Phys. Rev. **184**, 1848–1859 (1969). (A)

Two of the pedagogical reviews of this earlier period of anomaly studies can be found in:

116. **"Perturbation Theory of Anomalies,"** S. L. Adler, in *Lectures on Elementary Particles and Quantum Field Theory*, Proc. of 1970 Brandeis Summer Inst., edited by S. Deser, *et al.* (MIT Press, Cambridge, MA, 1971), 1–164. (A)
117. **"Field Theoretical Investigations in Current Algebra,"** R. Jackiw, in *Current Algebra and Anomalies* (Ref. 56). (A)

It is important to distinguish between anomalies of a global symmetry from those of a gauge symmetry. The former are physically relevant, while the latter will spoil the renormalizability of the gauge theory. For example, we need the anomaly in the (global) chiral symmetry of the strong interaction (in the presence of an external electromagnetic field) to understand the $\pi^0 \to \gamma\gamma$ decay, while the self-consistency of the renormalizable Weinberg–Salam electroweak theory depends on the absence of anomalies in the $SU(2)\times U(1)$ gauge symmetry (see Sec. III B). The conditions for the absence of anomalies in gauge theories were studied in the following papers:

118. **"An Anomaly-Free Version of Weinberg's Model,"** C. Bouchiat, J. Iliopoulos, and Ph. Meyer, Phys. Lett. **38B**, 519–523 (1972). (A)
119. **"Effect of Anomalies on Quasi-Renormalizable Theories,"** D. J. Gross and R. Jackiw, Phys. Rev. D **6**, 477–493 (1972). (A)
120. **"Gauge Theories without Anomalies,"** H. Georgi and S. L. Glashow, Phys. Rev. D **6**, 429–431 (1972). (A)

For a more recent review of this ever-expanding field, see Treiman *et al.* (Ref. 56) and:

121. **"Anomalies in Gauge Theories,"** A. Yu. Morozov, Sov. Phys. Usp. **29**, 993–1039 (1986). (A)

III. GAUGE THEORIES: APPLICATIONS

The successful formulation of gauge theories of the strong, weak, and electromagnetic interactions has been one of the major triumphs in particle physics. By now the gauge principle has become one of the most basic ingredients of the theoretical framework. We will now illustrate how the gauge principle is implemented in various types of interactions.

A. Quantum electrodynamics

The quantum theory of electrons interacting with photons, QED, is a gauge theory based on the $U(1)$ gauge group. Historically, this theory is the first relativistic quantum field theory ever formulated, and this is also where the basic idea of gauge invariance originates. For masterful accounts by one of its principal architects, we have, at the popular level:

122. **QED**, R. P. Feynman (Princeton U. P., Princeton, NJ, 1985). (E)

and, at a more advanced level, emphasizing calculational techniques (but still very brief):

123. **The Theory of Fundamental Processes**, R. P. Feynman (Benjamin, New York, 1961). (A)
124. **Quantum Electrodynamics**, R. P. Feynman (Benjamin, New York, 1962). (A)

The QED Lagrangian density as a function of the electron and photon fields, $\psi(x)$ and $A_\mu(x)$, is given by

$$\mathscr{L} = \bar{\psi} i\gamma^\mu(\partial_\mu - ieA_\mu)\psi - m\bar{\psi}\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu},$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{25}$$

This theory has two basic parameters, the electron mass m and the fine structure constant $\alpha = e^2/4\pi$. Since $\alpha \simeq \frac{1}{137}$ is small, one can, in principle, calculate perturbatively all the measurable quantities in terms of m and α to any desired accuracy. However, one encounters divergences in higher order loop diagrams. The theory of renormalization was originally developed to solve this problem in QED. Some of the basic QED renormalization papers are reprinted in Ref. 98. We have already cited some of the QED textbooks and monographs (Refs. 85–90) at the beginning of Sec. II E. For books containing more detailed calculations and comparison with experiments, see:

125. **The Quantum Theory of Radiation**, W. Heitler (Clarendon, Oxford, 1936). (A)
126. **Advanced Quantum Mechanics**, J. J. Sakurai (Addison-Wesley, Reading, MA, 1965). (A)
127. **Relativistic Quantum Mechanics**, J. D. Bjorken and S. Drell (McGraw-Hill, New York, 1965). (A)
128. **Quantum Electrodynamics**, A. Akhiezer and Y. Berestetskii (Wiley–Interscience, New York, 1965). (A)

The agreement between the theoretical predictions of QED and experimental results has been quite spectacular. For example, the magnetic moment of the electron (written here as half of the gyromagnetic ration: $\frac{1}{2}g_e$) is predicted to be

$$\mu_e^{(\text{th})} = (1\,001\,159\,652.48 \pm 0.23)\times 10^{-9}.$$

129. **"Eighth-order Anomalous Magnetic Moment of the Electron,"** T. Kinoshita and W. B. Lindquist, Phys. Rev. Lett. **47**, 1573–1576 (1981). (A)

The experimenal measurement gives

$$\mu_e^{(\text{exp})} = (1\,001\,159\,652.22 \pm 0.05)\times 10^{-9}.$$

130. **"New Comparison of the Positron and Electron g Factors,"** P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, Phys. Rev. Lett. **47**, 1679–1682 (1981). (A)

Thus, for practical purpose, the theory of QED is considered to be well understood and often it has been used as a tool for probing other interactions.

B. The SU(2)×U(1) electroweak gauge theory

The first successful application of the Yang–Mills theory is in the area of weak interactions. The non-Abelian $SU(2)\times U(1)$ gauge group provides a unified description of weak interactions and electromagnetism. It combines in one framework QED and the low-energy $V–A$ theory of charged current weak interactions. It is an example of the Yang–Mills theory with spontaneous symmetry breaking. The theory is renormalizable.

Much of the literature on weak interactions including unified gauge models and new flavor quantum numbers, etc. has been reviewed in previous Resource Letters:

131. **"Resource Letter WI-1: Weak Interactions,"** B. R. Holstein, Am. J.

Phys. **45**, 1033–1040 (1977). (E)

132. **"Resource Letter NP-1: New Particles,"** J. L. Rosner, Am. J. Phys. **48**, 90–103 (1980). (E)

133. **"Resource Letter Q-1: Quarks,"** O. W. Greenberg, Am. J. Phys. **50** 1074–1089 (1982). (E)

Since the end of the 1950s, it has been known that weak interactions can be described very well by the *V–A theory* with an effective Lagrangian very much like the one for β decays as originally suggested by Fermi,

$$\mathscr{L}_{\text{eff}} = (G_F/\sqrt{2})\, J_\lambda^\dagger J^\lambda, \tag{26}$$

where G_F is the Fermi coupling constant. If we restrict ourselves to the "first generation" of leptons and quarks,

$$\text{The first generation: } (e, \nu_e, d, u), \tag{27}$$

the charged *V–A* weak current can be written as

$$J^\lambda = \tfrac{1}{2}\bar{\nu}\gamma^\lambda(1-\gamma_5)e + \tfrac{1}{2}\bar{u}\gamma^\lambda(1-\gamma_5)d$$
$$= \bar{\nu}_L\gamma^\lambda e_L + \bar{u}_L\gamma^\lambda d_L, \tag{28}$$

where $e_L = \frac{1}{2}(1-\gamma_5)e$, etc. The *V–A* structure implies that only the left-handed helicity states appear in weak current. Compare this to the electromagnetic current

$$J^\lambda_{\text{em}} = -e\gamma^\lambda e + \tfrac{2}{3}\bar{u}\gamma^\lambda u - \tfrac{1}{3}\bar{d}\gamma^\lambda d, \tag{29}$$

and note the similarity of Eq. (26) to the electromagnetic amplitude $e^2 J^\lambda_{\text{em}} J^\mu_{\text{em}} (g_{\lambda\mu}/k^2)$ corresponding to photon exchange between two electromagnetic currents. One may look upon Eq. (26) as the effective interaction generated by a Yang–Mills theory with *massive* gauge bosons $W^\pm$ (hence, a coupling $gJ^\lambda W_\lambda$) mediating between the two weak currents. In the low-energy limit when the W propagator can be approximated by its mass factor, the Fermi constant can then be related to the gauge coupling g and the gauge boson mass M_W by

$$G_F \propto g^2/M_W^2. \tag{30}$$

If $W^\pm$ is to be identified with gauge bosons, what should the gauge group be? Schwinger, in the first attempt to unify QED and the weak interaction theory (before the advent of *V–A* theory), suggested that the photon (γ) and the $W^\pm$ form a triplet with respect to a new group SU(2):

134. **"A Theory of the Fundamental Interactions,"** J. Schwinger, Ann. Phys. **2**, 407–434 (1957). (A)

Once the weak current is known, as in Eq. (28), this group choice is actually unattainable. One can check that the weak charges T_+ and T_- (formed from the current J_λ and $J_\lambda^\dagger$) and the electromagnetic charge Q (from J^λ_{em}) do not have the SU(2) commutation relations. The simplest way out is to enlarge the gauge group to SU(2)×U(1) having SU(2) generators T_+, T_-, and $T_3 \equiv \frac{1}{2}[T_+, T_-]$, and a U(1) generator Y. Q is then some linear combination of T_3 and Y. In this way, one can make a representation assignment of all the fermions: For the first generation, the left-handed (ν_e, e) and (u, d) are SU(2) doublets and all the right-handed states are SU(2) singlets. The U(1) charges are then "put-in-by-hand" so that their electromagnetic charges come out correctly. The proposal that the weak gauge group should be SU(2)×U(1) was first made by Glashow:

135. **"Partial-Symmetries of Weak Interactions,"** S. L. Glashow, Nucl. Phys. **22**, 579–588 (1961). (A)

However, the gauge boson masses were inserted by hand and thus the gauge symmetry was explicitly broken. It turns out that such a theory of massive vector bosons is not renormalizable. In 1967, the crucial step was taken by Weinberg, and independently by Salam, to generate the gauge boson masses in this SU(2)×U(1) model via spontaneous symmetry breaking by the Higgs mechanism as discussed in Sec. II C. The simplest way to implement this is to have a doublet of complex scalars ϕ, which develops a vacuum expectation value $\langle\phi\rangle \neq 0$. This breaks the gauge symmetry down to the electromagnetic U(1):

$$\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_{\text{em}}. \tag{31}$$

Of the four independent scalar components, three are combined with the original massless gauge fields (having only transverse states) to form massive vector particles $W^\pm$ and Z^0 (now having longitudinal states as well) with $M_{W,Z} \propto \langle\phi\rangle$, and one linear combination, the photon, is still massless. The remaining scalar survives as a massive particle, the *Higgs boson*.

*136. **"A Model of Leptons,"** S. Weinberg, Phys. Rev. Lett. **19**, 1264–1266 (1967). (A)

*137. **"Weak and Electromagnetic Interactions,"** A. Salam, in *Elementary Particle Physics*, Nobel Symp. No. 8, edited by N. Svarthom (Almqvist and Wiksell, Stockholm, 1968), 367–377. (A)

Weinberg and Salam speculated that such a theory was renormalizable. This conjecture was vindicated when 't Hooft proved in 1971 that spontaneously broken gauge theories were renormalizable (Ref. 104). The achievement of 't Hooft immediately placed the Weinberg–Salam model in the center stage as the leading candidate field theory of weak interactions; in a broader context, it transformed the field of elementary particle theory. For a concise history of gauge theory development, particularly that of the electroweak unification of Glashow, Weinberg, and Salam, see:

138. **"The 1979 Nobel Prize in Physics,"** S. Coleman, Science **206**, 1290–1292 (1979). (E)

The original proposals of Weinberg and Salam pertained only to the leptons. A proper theory of hadronic weak interactions was not possible unless, in more modern parlance, the second fermion generation is completed with another quark, the charm quark. Just as the mu-lepton doublet (μ, ν_μ) is the second generation extension of (e, ν_e), lepton–quark symmetry would require that the strange quark be paired with a new charm quark (s,c) which extends the (d,u) doublet:

139. **"Elementary Particles and *SU*(4),"** J. D. Bjorken and S. L. Glashow, Phys. Lett. **11B**, 255–257 (1964). (A)

But not until 1970 was the key physics rationale provided by Glashow, Iliopoulos, and Maiani who showed that any sensible weak interaction theory must have this new charm quark in order to suppress to an acceptable level the induced strangeness-changing neutral current effects such as $K^0 \to \mu^+\mu^-$:

140. **"Weak Interactions with Lepton–Hadron Symmetry,"** S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D **2**, 1285–1292 (1970). (A)

In terms of the gauge theory of weak interactions, the introduction of charm means that

$$\text{The second generation: } (\mu, \nu_\mu, s, c) \tag{32}$$

has the identical gauge couplings as the first generation of (27):

141. **"Effects of a Neutral Intermediate Boson in Semileptonic Processes,"** S. Weinberg, Phys. Rev. D **8**, 1412–1417 (1972). (A)

Hadrons containing charm quark were discovered in 1974. In fact, we now know that there must be at least three generations of leptons and quarks. Besides (27) and (32), we also have

The third generation: (τ,ν_τ,b,t). (33)

(As of this writing, there is still no firm direct experimental evidence for the top quark t.) For references of these new particles, see Rosner (Ref. 132).

This repetitive fermion structure in the theory explains the presence of the mysterious mixing angle in charged currents, the "*Cabibbo angle*":

142. "**Unitary Symmetry and Leptonic Decays,**" N. Cabibbo, Phys. Rev. Lett. **10**, 531–533 (1963). (A)

and it also provides a possible theory of CP violation:

143. "**CP-Violation in the Renormalizable Theory of Weak Interactions,**" M. Kobayashi and K. Maskawa, Prog. Theor. Phys. **49**, 652–657 (1973). (A)

However, the underlying reason for this repetitive fermion generation structure is still not understood. In fact, the theory with just one generation is self-consistent by itself. For example, while the various fermion representations will give rise to chiral anomalies that will destroy the renormalizability of the theory (see Sec. II E), the lepton and quark anomaly contributions miraculously cancel within each generation (Refs. 117 and 118).

The gauge group $SU(2)\times U(1)$ has $3+1$ gauge bosons. Besides the photon field A and the intermediate vector bosons $W^{\pm}$ mediating the charged current weak processes such as $\nu+p\rightarrow e+\cdots$, there is also the gauge boson Z, which transmits a set of hitherto unobserved *neutral current processes* like $\nu+p\rightarrow\nu+\cdots$. In 1973, the first experimental evidence of a neutral current event was reported by the bubble chamber group Gargamelle working at CERN. One event of $\nu_\mu+e\rightarrow\nu_\mu+e$ and hundreds of $\nu_\mu+N\rightarrow\nu_\mu+\cdots$.

144. "**Search for Elastic Muon-Neutrino Electron Scattering,**" F. J. Hasert *et al.*, Phys. Lett. **46B**, 121–124 (1973).

145. "**Observation of Neutrino-like Interactions Without Muon or Electron in the Gargamelle Neutrino Experiments,**" F. J. Hasert *et al.*, Phys. Lett. **46B**, 138–140 (1973).

Many other neutral current experiments were performed after its discovery: νN-, νe-scatterings, forward–backward asymmetry in $e^+e^-\rightarrow\mu^+\mu^-$, atomic parity violation, and polarized electron deuteron experiments, etc. All these processes can be predicted in this theory in terms of one parameter θ_W, the *Weinberg angle*, where $\tan\theta_W=g'/g$, g' and g being the U(1) and SU(2) gauge couplings, respectively. The fact that this large amount of data can all be fitted with

$$\sin^2\theta_W\simeq 0.23 \tag{34}$$

is strong evidence for the correctness of the $SU(2)\times U(1)$ theory. For the most recent result, see

146. "**Comprehensive Analysis of Data Pertaining to the Weak Neutral Current and the Intermediate Vector–Boson Masses,**" U. Amaldi *et al.*, Phys. Rev. D **36**, 1385–1407 (1987). (A)

The most basic and distinctive feature of a gauge theory of weak interactions is the existence of a set of massive gauge bosons. In the Weinberg–Salam theory, Eq. (30) has the precise form of

$$G_F/\sqrt{2}=g^2/8M_W^2. \tag{35}$$

For the simplest Higgs representation, we also have $M_Z=M_W/\cos\theta_W$. Thus for the value of θ_W as given in Eq. (34), the theory predicts, after higher-order effects are included, M_W and M_Z to be 82 and 93 GeV, respectively. In 1983, the UA1 group at CERN (Rubbia *et al.*) made these discoveries of the W and Z with properties precisely as predicted by the $SU(2)\times U(1)$ theory.

147. "**Experimental Observations of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s}=540$ GeV,**" G. Arnison *et al.*, Phys. Lett. **122B**, 103–106 (1983). (A)

148. "**Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at CERN SPS Collider,**" G. Arnison *et al.*, Phys. Lett. **126B**, 398–410 (1983). (A)

The electroweak theory is basically complete except for a few loose ends. The top quark still awaits discovery. This is anticipated by just about everyone as the renormalizability of the theory requires this completion of the third generation (33). Since we do not really understand the origin of the fermion generation structure in the theory, the existence of higher generations may also be possible. The Weinberg–Salam model does not predict the mass of the Higgs boson, and there is no experimental evidence for its existence. It is anticipated that high-energy experiments that will be possible with new generations of accelerators in the next two decades will shed light on this important area.

In addition to the nontechnical articles, the textbooks, and monographs cited in Sec. I (Refs. 13–41), we also recommend the following discussions of the electroweak gauge theory:

149. **Weak Interactions**, D. Bailin (Hilger, Bristol, 1982), 2nd ed. (A)

150. **Weak Interactions of Leptons and Quarks**, E. Commins and P. H. Bucksbaum (Cambridge U. P., London, 1983). (A)

151. **Weak Interactions and Modern Particle Theory**, H. Georgi (Benjamin/Cummings, Menlo Park, CA, 1984). (A)

C. Quantum chromodynamics (QCD)

The key steps to the discovery of QCD as the candidate quantum field theory of strong interaction are the proof of renormalizability of the Yang–Mills theory (see Sec. II E) and the establishment of the properties of asymptotic freedom (see Sec. II D). The success of the parton model interpretation of the scaling phenomenon in deep inelastic electron–nucleon scatterings means that quarks must be tightly bound inside the hadrons (as quarks are not freely produced) yet at small distances corresponding to large momentum transfers of the deep inelastic scatterings they behave as if they were free:

152. **An Introduction to Quarks and Partons**, F. E. Close (Academic, New York, 1979).

Thus in order for a quantum field theory to describe a quark theory of hadrons it must have the property of asymptotic freedom. This singles out non-Abelian gauge theory, see Gross and Wilczek (Ref. 76), Politzer (Ref. 77), and Coleman and Gross (Ref. 78). At long distances, the theory should offer the possibility for a confinement mechanism ("infrared slavery"). This implies that the non-Abelian gauge symmetry must not be broken and the infrared divergence due to the massless gauge bosons is such that all gauge symmetry charges are confined.

Which symmetry of the quark model should be gauged? Recall that the simple quark model as envisioned in the early 1960s has the serious problem that the wavefunctions of the flavor-SU(3) decuplet baryons seem to violate the connection between spin and statistics. Take the example of N^{*++}. Since it is a ground state for the system of three u quarks, the spatial wavefunction has zero total angular momentum and is totally symmetric. But N^{*++} has spin-3/2 and the spins of all u quarks must be lined up so that the spin wavefunction is also totally symmetric. Consequently, the overall wavefunction is symmetric with respect to the interchange of any pairs of constituent quarks. This violates Fermi–Dirac statistics since the quark is a spin-1/2 fermion. The way out of this and other related difficulties gradually emerges: Greenberg first suggested *para-Fermi statistics of order 3* for the quarks:

153. **"Spin and Unitary-Spin Independence in a Paraquark Model of Baryons and Mesons,"** O. W. Greenberg, Phys. Rev. Lett. **13**, 598–602 (1964). (A)

It is simpler to describe this new degree of freedom for quarks by saying that each flavor quark carry three "*colors*." They transform as fundamental **3** representation under the *color*-SU(3) group (as opposed to the original *flavor*-SU_3 of Gell-Mann and Ne'eman) while all observed hadrons are color singlets. Thus $N^{*++} \sim \epsilon_{\alpha\beta\gamma} u_\alpha u_\beta u_\gamma$, all the color indices α, β, and γ contracted with the Levi–Civita tensor to form a color singlet, has now a totally antisymmetric wavefunction to satisfy the Fermi–Dirac statistics.

154. **"Three-Triplet Model with Double $SU(3)$ Symmetry,"** M. Y. Han and Y. Nambu, Phys. Rev. **139B**, 1006–1010 (1965). (A)

155. **"A Systematic of Hadrons in Subnuclear Physics,"** Y. Nambu, in *Prelude in Theoretical Physics*, edited by A. de Shalit *et al.* (North-Holland, Amsterdam, 1966), pp. 133–142. (A)

The idea of fractionally charged quarks with exact SU(3) color degeneracy, with hadrons all being color singlets (and the terminology "color") were not generally accepted until they were advocated by Gell-Mann and his coworkers in 1972.

156. **"Quarks,"** M. Gell-Mann, in *Elementary Particle Physics*, edited by P. Urban (Springer, Vienna, 1972); Acta. Phys. Austriaca Suppl. **9**, 733–761 (1972). (A)

157. **"Light Cone Current Algebra, π^0 Decay and e^+e^- Annihilation,"** W. A. Bardeen, H. Fritzsch, and M. Gell-Mann, in *Scale and Conformal Symmetry in Hadron Physics*, edited by R. Gatto (Wiley, New York, 1973), 139–151. (A)

The title of the last paper refers to the fact that, in the cases of the anomaly calculation of the $\pi^0 \to \gamma\gamma$ decay rate (Refs. 113 and 114) and the parton model calculation of e^+e^- annihilation into hadrons, agreement between theory and experiment depends critically on quarks having this hidden color degree of freedom. For a historical review of the color concept development, see

158. **"Color Models of Hadrons,"** O. W. Greenberg and C. A. Nelson, Phys. Rep. **32C**, 69–121 (1977). (A)

Furthermore, since we also need to assume that only color singlets are observable, it suggests that the forces between the color quarks must be color dependent. All this leads to the idea that it is the color symmetry of the quark model that should be gauged, and the basic strong interaction should be described by a color-SU(3) Yang–Mills theory where each flavor of quarks transforms as the fundamental triplet representation. This, together with the requirement that the theory be renormalizable, (almost) completely fixes the form of the QCD Lagrangian:

$$\mathscr{L}_{\text{QCD}} = -\tfrac{1}{2}\,\text{tr}\,G_{\mu\nu}G^{\mu\nu} + \sum_{k=1}^{n_f} \bar{q}_k (i\gamma^\mu D_\mu - m_k) q_k,$$

where

$$\begin{aligned} G_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu - ig_s[A_\mu, A_\nu] \\ D_\mu q_k &= (\partial_\mu - ig_s A_\mu) q_k \\ A_\mu &= \sum_{a=1}^{8} \frac{A_\mu^a \lambda^a}{2} \end{aligned} \qquad (36)$$

and g_s is the (strong) gauge coupling constant. The q_k's are the quark fields with flavor index $k = u,d,s,c,b$, and t. Summation over their color indices (which are not displayed) is understood. The λ^a's are the Gell-Mann $SU(3)$ matrices in color space and the eight *gluons* A_μ^a are the confined massless gauge particles. This elegant theory of the strong interaction was proposed almost immediately after the discovery of asymptotic freedom, see Gross and Wilczek (Ref. 76), Weinberg (Ref. 83), and Fritzsch, Gell-Mann, and Leutwyler (Ref. 84).

Because of the asymptotic freedom, calculations of short-distance (i.e., high-momentum transfer Q^2) physics are relatively tractable in QCD. To the first approximation QCD reproduces parton model results (scaling). The renormalization group improved perturbation theory also specifies the approach of this limit (scaling violation). All the high Q^2 experiments are in semiquantitative agreement with these calculations. For review, see Gross (Ref. 82) and

159. **"Asymptotic Freedom: An Approach to the Strong Interaction,"** H. D. Politzer, Phys. Rep. **14**, 129–180 (1974). (A)

160. **"Perturbative Quantum Chromodynamics,"** E. Reya, Phys. Rep. **69C**, 195–334 (1981). (A)

161. **"Partons in Quantum Chromodynamics,"** G. Altarelli, Phys. Rep. **81C**, 1–130 (1982). (A)

162. **"First Lap in QCD,"** F. Close, Phys. Scripta **25**, 86 (1982). (A)

The observation of jet events in high-energy hadron productions (i.e., hadrons flowing in cones of small width) constitute perhaps the most direct evidence for the existence of quarks and gluons. For example in e^+e^- annihilations into hadrons we have the two-jet structure corresponding to the quark and antiquark pairs:

163. **"Evidence for Jet Structure in Hadron Production e^+e^- Annihilation,"** G. Hanson *et al.*, Phys. Rev. Lett. **35**, 1609–1612 (1975). (A)

and the three-jet events of $q\bar{q}$ and a gluon:

164. **"Evidence for Gluon Bremsstrahlung in e^+e^- Annihilations at High Energies,"** Ch. Berger *et al.*, Phys. Lett. **86B**, 418–425 (1979). (A)

The bulk of strong interaction effects involves long-distance physics: confinement, hadron masses, and quark and gluon fragmentation into hadrons, etc. At low Q^2 the effective coupling is larger than unity, one needs some nonperturbative method to study QCD. The most powerful technique is *lattice-QCD* invented by Wilson.

165. **"Confinement of Quarks,"** K. G. Wilson, Phys. Rev. D **10**, 2445–2459 (1974).

166. **"Quarks and Strings on a Lattice,"** K. G. Wilson, in *New Phenomena in Subnuclear Physics*, Proc. 1975 Int. Sch. Subnucl. Phys. 'Ettore Majorana', edited by A. Zichichi (Plenum, New York, 1977). (A)

In lattice gauge theory space-time continuum is discretized. This provides a natural cutoff of length scale by the lattice spacing. With finite lattice size, it is possible to study various physical quantities in the path integral formulation by computer simulation based on the Monte Carlo method. For general reviews, see:

167. **"An Introduction to Lattice Gauge Theory and Spin System,"** J. Kogut, Rev. Mod. Phys. **51**, 659–713 (1979). (A)

168. **Quarks, Gluons and Lattice**, M. Creutz (Cambridge U. P., London, 1983). (A)

169. **Lattice Gauge Theories and Monte Carlo Simulations**, C. Rebbi (World Scientific, Singapore, 1983). (A)

In addition to the textbooks and monographs cited in the Introduction, we have these self-contained reviews of QCD:

170. **"Quantum Chromodynamics,"** W. Marciano and H. Pagels, Phys. Rep. **36C**, 137–276 (1978).

171. **"Quantum Chromodynamics: The Modern Theory of the Strong Interaction,"** F. Wilczek, Ann. Rev. Nucl. Part. Sci. **32**, 177–209 (1982).

172. **Quantum Chromodynamics–An Introduction to the Theory of Quarks and Gluons**, F. J. Yuduráin (Springer-Verlag, Berlin, 1983).

D. Grand unified theories (GUTs)

The Weinberg–Salam electroweak theory and QCD theory of strong interaction are collectively referred to as the "*standard model*" of elementary particle physics. This SU(3)×SU(2)×U(1) gauge theory is compatible with all well-established experimental results. Still there is plenty of motivation to construct a more unified theory. In this subsection, we briefly discuss the first step in going beyond the standard model: the study of "*grand unified theories*" (GUTs) in which a simple gauge group with one gauge coupling describes both the strong and the electroweak interactions.

The basic scheme is to have a grand unification group G that is spontaneously broken down, in one or several steps, to the standard model gauge group $G\rightarrow SU(3)_c\times SU(2)_L\times U(1)_Y\rightarrow SU(3)_c\times U(1)_{em}$. If the spontaneous symmetry breakings are through the Higgs mechanism (Sec. II C), we need at least two sets of Higgs scalars developing vacuum expectation values $\langle\phi_{GUT}\rangle \gg \langle\phi_{SM}\rangle$ so that the gauge bosons of G that do not belong to the subgroup SU(3)×SU(2)×U(1) acquire masses $M_X\propto\langle\phi_{GUT}\rangle$ much larger than the weak intermediate vector boson masses $M_W\propto\langle\phi_{SM}\rangle$.

Georgi, Quinn, and Weinberg have shown that this picture can account for the observed differences in the SU(3), SU(2), and U(1) couplings g_s, g, and g', because these are effective couplings for energy scales much lower than the GUTs scale M_X.

*173. **"Hierarchy of Interactions in Unified Gauge Theories,"** H. Georgi, H. R. Quinn, and S. Weinberg, Phys. Rev. Lett. **33**, 451–454 (1974). (A)

If we make the simplifying hypothesis, the so-called "*desert assumption*," that there is essentially no new particle threshold lying between M_W and M_X, then we have the following picture of coupling unification: Three smooth curves of the effective couplings as functions of energy converge to one point at M_X. Since their energy dependencies, being determined by the renormalization group equations of their respective gauge groups, are only logarithmic, and since g_s, g, and g' are quite different at present energies, the unification scale M_X turns out to be extremely high, $\gtrsim 10^{14}$ GeV. The requirement that they meet at one point implies a nontrivial constraint among g_s, g, and g'. Taking g_s and the electromagnetic coupling e as known, this leads to the value of $\sin^2\theta_W\simeq 0.23$ in general agreement with neutral current results of Eq. (34).

Early attempts to put quarks and leptons in the same gauge group multiplets were made in:

174. **"Unified Lepton–Hadron Symmetry and a Gauge Theory of the Basic Interactions,"** J. C. Pati and A. Salam, Phys. Rev. D. **8**, 1240–1251 (1973). (A)

Georgi and Glashow then showed, in a very influential paper,

*175. **"Unity of All Elementary Particle Forces,"** H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438–441 (1974). (A)

that the smallest grand unification group containing SU(3)×SU(2)×U(1) as a subgroup and having complex representations (to allow the two fermion helicity states to have different gauge couplings as in the standard model) is SU(5). In fact, the **5*** and **10** representations of SU(5) have exactly the right SU(3)×SU(2)×U(1) decomposition to account for each generation of fermions. For the first generation,

$$\mathbf{5}^*:\left[\begin{pmatrix}\nu\\ e\end{pmatrix},\bar{d}^{\alpha}\right]\quad \mathbf{10}:\left[\begin{pmatrix}u^{\alpha}\\ d^{\alpha}\end{pmatrix},\bar{u}^{\alpha},e^{+}\right],$$

$\alpha=1,2,3$ is the color index. The unification group being simple; *charge quantization* is obtained. Indeed some light is shed on the fractional $\frac{1}{3}$ charge assignment of quarks; they are essentially the lepton charge divided by the number of colors. Remarkably also, anomalies coming from the **5*** and **10** representations mutually cancel.

SU(5) is the smallest possible grand unified group. The natural extension to higher-rank groups are the orthogonal group SO(10).

176. **"The State of the Art-Gauge Theories,"** H. Georgi, in *Particles and Fields—1974*, Proc. Mtg. APS Div. Part. Fields, Williamsburg, VA, edited by C. E. Carlson (AIP, New York, 1975), 575–584. (A)

177. **"Unified Interactions of Leptons and Hadrons,"** H. Fritzsch and P. Minkowski, Ann. Phys. **93**, 193–265 (1975). (A)

and the exceptional group E(6):

178. **"A Universal Gauge Theory Model Based on E_6,"** F. Gürsey, P. Ramond, and P. Sikivie, Phys. Lett. **60B**, 177–180 (1978). (A)

179. **"Quark–Lepton Symmetry and Mass Scales in an E_6 Unified Gauge Model,"** Y. Aichiman and B. Stech, Phys. Lett. **77B**, 389–396 (1978). (A)

One of the attractive features of such GUTs is that fermions in each generation can belong to one irreducible representation [in the case of $E(6)$ it is the smallest (fundamental) representation], and the miraculous anomaly cancellations in the standard model and SU(5) are explained [because these embedding groups SO(10) and $E(6)$ are "*anomaly-free*" (see Ref. 120)]. However, the price one pays is that we must have, in each generation, fermions beyond the presently known leptons and quarks. For a detailed survey of all possible grand unification groups, see

180. **"Color Embeddings, Charge Assignments, and Proton Stability in Unified Gauge Theories,"** M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. **50**, 721–744 (1978). (A)

All GUTs predict proton decay. When the "desert assumption" is realized in the SU(5) model with the simplest Higgs structure, this leads to definite predictions of the decay rate and branching ratios. These predictions have been contradicted by experiments. However, there are compelling physical reasons to populate the desert (supersymmetry to solve the "*hierarchy problem*") that will significantly increase the proton decay rate without drastic change of the $\sin^2\theta_W$ prediction.

Comprehensive review of grand unified theories can be found in the following references:

181. **"Grand Unified Theories and Proton Decay,"** P. Langacker, Phys. Rep. **72C**, 185–385 (1981). (A)

182. **Grand Unification**, G. Ross (Benjamin/Cummings, Reading, MA, 1985). (A)

183. **Unification and Supersymmetry**, R. Mohapatra (Springer-Verlag, Berlin, 1986). (A)

For a review and a collection of important GUTs papers, see

184. **Unity of Forces in the Universe**, A. Zee (World Scientific, Singapore, 1982). (A)

ACKNOWLEDGMENTS

This work was supported in part by the United States Department of Energy, National Science Foundation

(PHY-8605700 and PHY-8620266), and the University of Missouri Weldon Spring Fund. Part of this work was performed while one of us (T.P.C.) was a participant at the Aspen Center for Physics, and while we were visiting members of The Institute for Advanced Study. The hospitality of these institutions is gratefully acknowledged.

[a] Permanent address.

The amazing many colored relativity engine

N. David Mermin

Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853-2501

(Received 17 August 1987; accepted for publication 15 October 1987)

A simple microcomputer program is described that makes beginning students think relativistically about elementary space and time measurements. The program allows such students to perform their own *gedanken* experiments to learn the consequences a clock synchronization convention has for measurements of the lengths or speeds of moving objects and the rates of moving clocks. To use the program requires no knowledge of physics, algebra, or geometry.

I. AN ENGINE FOR TEACHING RELATIVITY

I describe below a simple method for teaching the elementary (special) relativistic properties of space and time measurements. This approach was developed and refined during several teachings of a "general education" course on special relativity for nonscientists. Such students are not at home with elementary algebraic manipulations and a direct derivation of the Lorentz transformation along conventional lines is beyond their conceptual powers (nor would they be able to use the equations if they had them). Even a step by step derivation of length contraction, time dilation, and the relativity of simultaneity requires a sophistication with algebraic formalism beyond the attainments of most of these students, and they are entirely distracted from genuine physical and philosophical subtleties by their marginally successful efforts to cope with the mathematics.

One can remove much of this irrelevant but overpowering obscurity by replacing algebra with arithmetic. To make many of the conceptual points it is not necessary to derive everything for general values of v/c. Almost everything is illustrated just as well by special cases, for example $v/c = \frac{3}{5}$. Many students who strain mightily over $(1 - v/c)(1 + v/c) = 1 - v^2/c^2$ will swallow $\frac{2}{5}\times\frac{8}{5} = \frac{16}{25}$ with relative ease.

But the arithmetical approach cannot avoid all algebra, and for some students even arithmetic distracts from the real issues. Stimulated by this chronic source of frustration, I was led to design what might be characterized as a hands-on *gedanken* experiment. In this *gedanken* laboratory length contraction, time dilation, relativity of simultaneity, velocity addition, and the like are extracted directly from measurements made with one simulated set of instruments to determine the properties of another such set. No calculations are required beyond the arithmetical manipulation of data. The mutual consistency of all measurements is self-evident. And all the major relativistic effects are directly revealed.

My original idea, in the late 1960s, was to build this pedagogical device into something like a relativistic slide rule—an analog computer, with various sets of gear-linked counter-rotating concentric cylinders representing moving meter sticks and the readings on attached clocks. I imagined the whole thing in brass on a mahogany base, driven by a large ebony-handled crank: a Relativity Engine.

They laughed when I brought them drawings. Nobody would make it for me. So I had to make do with sheets of paper schematically depicting various slices of this gorgeous space-time Relativity Engine.

I got by with pieces of paper throughout the next decade and a half until I noticed that slide rules had vanished, and the campus was suddenly teeming with little computers. The time had come to simulate my Engine on the display screen; students would be able to turn the crank at any of hundreds of conveniently located keyboards. Mahogany and brass were out, but color was in.

This article describes my digital Relativity Engine.

Section II gives a general view of the Engine. It is written with teachers of physics in mind, but it should be intelligible to students with some prior acquaintance with special relativity who might also enjoy playing with the Engine.

Section III is an introduction to the Engine for students who have no acquaintance whatever with special relativity. It is in effect a user's manual. It will also provide teachers with a more detailed picture of how the Engine operates and the pedagogical possibilities it gives rise to. Nowhere in Sec. III do I attempt to put the Engine into the context of an introductory course in special relativity: There are too many different ways one might try, and it seemed to me this was really a matter for the judgment of the individual teacher.

Some suggestions along these lines are given in Sec. IV A, which is again addressed to teachers. Section IV B gives a brief formal analysis of the Engine, which might help those wishing to design Engines of their own with parameters different from mine. Section IV C discusses a few technical features of my computer program and tells interested readers how to get a copy for themselves.

