
Over the past fifteen years a consensus has developed among elementary particle physicists that most, if not all, interactions between the fundamental particles of nature are described by gauge theories. During the last decade it has become clear that a gauge theory, at least in its classical aspects, is in fact a theory of connections and curvatures in principal bundles and associated vector bundles over (suitably compactified) spacetime manifolds, and this recognition has led to some nontrivial mathematical results. The book by Bleecker is a timely introduction to the differential geometry and variational principles of classical gauge theories.

Before proceeding to a brief review of the book, I would like to make a general remark on the strange interactions between mathematics and physics, and what Wigner called the "unreasonable effectiveness of mathematics" in modern physics, [Wigner, 1960]. One must begin to wonder why, from time to time, the disciplines of theoretical physics and modern mathematics drift apart, seem to develop in relative independence from each other, and then suddenly find domains of overlap and cross-fertilization. This happened before in this century (I have in mind the parallel developments of quantum mechanics, Hilbert space theory, and the theory of group representations, the developments of operator algebra theory and algebraic quantum field theory and statistical mechanics, inverse scattering methods and solitons, and most intriguing of all, the pervasiveness of modern differential geometry and topology in developments in classical mechanics, general relativity, and now gauge theories; another recent related development is "supersymmetry" also known as the theory of graded Lie algebras, which is rapidly making contacts with gauge theories both in its physical, and in its mathematical aspects). One can only wonder about the deeper reasons of these developments and look forward to further cross-fertilization.

The term gauge invariance (Eichinvarianz in the original) together with the fundamental idea of a gauge theory was introduced by Hermann [Weyl, 1918] in an attempt to incorporate electromagnetism into a unified theory of gravitation and electromagnetism. His idea of "gauging" consisted in extending Einstein's principle of relativity by assuming that the scale of length can vary smoothly from point to point in spacetime. This led to the introduction of a one-form $A = A_\mu dx^\mu$ (identified with the electromagnetic four-potential) and
the corresponding “curvature” $F = \frac{1}{2} F_{\mu \nu} dx^\mu \wedge dx^\nu$ (identified with the electromagnetic field strength tensor $F_{\mu \nu}$) satisfying the Maxwell equations. Although this beautiful theory did not survive Einstein’s penetrating criticism (cf. the remarks by Weyl in [Weyl, 1955]), it was seminal to the development of gauge theories as we now know them, once it was recognized that the quantity to be gauged (i.e., made space-time dependent) was not the conformal factor in the metric, but rather the phase of the Schrödinger wave function (or its nonabelian equivalent—the parameters of some nonabelian Lie group describing a particle symmetry). This recognition evolved over a decade from the original proposal through some early remarks by [Schrödinger, 1922], before the invention of quantum mechanics (at about the time de Broglie introduced his waves into quantum theory), followed by work of [Fock, 1927] and [London, 1927], and was formulated in definitive form by [Weyl, 1929].

Before explaining the meaning of gauge invariance it is important to understand how fields enter into the description of elementary particles. At the classical level only two physical fields are known: the electromagnetic field, and the gravitational field. However, it is convenient to associate a field to each known particle in the following way: in quantum theory one may think of particles being created out of the vacuum into states with definite quantum numbers (energy, momentum, helicity = projection of spin onto momentum, charge, etc.). This is described by a creation operator acting on an appropriate Hilbert space of states. The creation-annihilation operators, multiplied by functions describing the states of the particles are then combined into “operator-valued fields” (more correctly, operator-valued distributions). These operator-valued fields are then subjected to various transformations, such as rotations, translations, Lorentz boosts, and gauge transformations, to which we turn our attention now.

Weyl’s gauge principle was based on the following idea: since the Schrödinger (or Dirac) wave field $\psi$ describing an electron may be multiplied by a phase factor $\exp(i \chi)$, i.e., subjected to a transformation belonging to the group $U(1)$, without any observable effect (since all observables involve the product of $\psi$ and its complex conjugate $\bar{\psi}$, which is multiplied by $\exp(-i \chi)$), observers located at different points in space should be allowed to choose different phase factors. In other words, the action of the group $U(1)$ on the function $\psi(x)$ is replaced by the action on it of the infinite-dimensional group whose elements are the point-dependent phase factors $\exp(i \chi(x))$ (in present terminology the action of the automorphism group of a principal $U(1)$-bundle on the sections of an associated line-bundle, the latter being represented by the wave functions $\psi(x) : \psi(x) \mapsto \exp(i \chi(x)) \cdot \psi(x)$).

However, since multiplication by $\exp(i \chi(x))$ does not commute with partial differentiation with respect to coordinates and time, the transformed wave function no longer satisfies the Schrödinger (or Dirac) equation, i.e., these equations are not invariant under such gauge transformations “of the second kind”. In order to restore invariance the partial derivatives (or exterior differentials) have to be replaced by “covariant derivatives” $\nabla_\mu = \partial_\mu + ieA_\mu$, where $A$ is again the electromagnetic potential and $e$ is the electron charge. The
commutator of two covariant derivatives (or the exterior differential of the one-form \( A \alpha \text{d}x^\alpha \)) is easily identified with the electromagnetic field strength tensor (or two-form) \( F \), which satisfies the Maxwell equation \( \text{d}F = 0 \). The other set of Maxwell equations follows, e.g., from the Schwarzschild action principle and identifies the electromagnetic four-current as the Weyl divergence of \( F \): \( \text{d}^*F = \ast J \). This \( U(1) \)-gauge theory became in the 1930s a standard tool of electrodynamics, and its definitive form was given by Pauli in his 1939 Solvay report, [Pauli, 1941].

The more interesting nonabelian gauge theories made their first sporadic appearance in an obscure paper by Oscar Klein [1938] (a paper which went unnoticed by the physics community and was forgotten even by its author, to surface only in the 1970s, when gauge theories were honored by three Nobel prizes). The first widely known nonabelian gauge theory, in which the group which was “gauged” was the isospin group \( SU(2) \) is due to Chen Ning Yang and Robert L. Mills, [Yang-Mills, 1954] (a similar theory seems to have been proposed independently by [Shaw, 1954]). At that time it was believed that the forces between the nucleons (protons and neutrons) were mediated by the exchange of pions, and the interaction was to a good approximation invariant under the group \( SU(2) \), with the proton and neutron forming a fundamental doublet (isospinor) of this group and the three charge states of the pion \((+1,0,-1)\) being a triplet in the adjoint representation (isovector). Fermi and Yang had considered a model in which the spin 1, isospin 1 pions were bound states of the more fundamental spin 1/2, isospin 1/2 nucleons and antinucleons, and it was natural to search for a gauge mechanism responsible for the binding of these fermions into boson states.

The Yang-Mills paper, although written for the specific example of \( SU(2) \), contained all the elements of contemporary nonabelian gauge theory, except for its explicit geometric interpretation in terms of principal bundles. However, a close reading of this paper reveals the deep geometric insight of its authors. It was pointed out that the analog of the electromagnetic four-potential is a covariant four-vector called the gauge potential, now recognized as the components of (the pullback of) a connection one-form. This one-form is subject to gauge transformations “of the second kind”. In modern notation, denoting the matrix-valued one-form by \( A \), these transformations are \( A \mapsto \text{Ad}(g^{-1})A + g^{-1}\text{d}g \), where \( g(x) \) is a smooth function on spacetime with values in the group \( SU(2) \). Such a “matrix-valued function on spacetime” represents a gauge transformation, i.e., an automorphism of the trivial principal bundle with spacetime as base and \( SU(2) \) as structure group. The curvature two-form \( F = \text{d}A + A \wedge A \), known as the Yang-Mills field strength, is subject to an adjoint transformation \( F \mapsto \text{Ad}(g^{-1})F \) under a gauge automorphism, and satisfies the Bianchi identity \( DF + A \wedge F = 0 \) (the “homogeneous Yang-Mills equation”). The whole set of equations is conformally invariant—where physics corresponds to the masslessness of the associated spin 1 (vector) particles. The coupling of the Yang-Mills field to its sources—the spin 1/2 particles—is achieved by replacing derivatives by covariant derivatives in the Dirac equations and noting that the covariant differential of the Hodge-dual of \( F \) is the fermionic current (this can be derived from a variational principle, just
like in electrodynamics). This brief description touches only the classical (nonquantum) aspects of the Yang-Mills theory since these are the only aspects treated in Bleecker's book. The Yang-Mills paper also deals with quantization, which we shall not discuss here.

It was only natural that the Yang-Mills theory should be amplified and extended to general Lie groups [Utiyama, 1956], [Mayer, 1956, 1959], and that the idea should also be applied to general relativity as a gauge theory of the Lorentz group (Utiyama, loc. cit. [Kibble, 1961], [Thirring, 1960]). The discovery in the late 1950s of the spin 1 resonances $\rho$ and $\omega$, which were considered candidates for Yang-Mills particles (e.g., [Sakurai, 1961], where a gauge theory of strong interactions is discussed), which ultimately led to the introduction in 1961 of $SU(3)$ as the group for strong interactions (known as the eightfold way, or flavor-$SU(3)$) by [Gell-Mann and Ne'eman, 1964], as well as the intermediate-boson model for the weak interactions and attempts to unify weak and electromagnetic interactions ([Schwinger, 1957], [Mayer, 1958], [Bludman, 1958], [Salam and Ward, 1959], [Gell-Mann and Glashow, 1961]), led to renewed interest in nonabelian gauge theories. However, conformal invariance required that the Yang-Mills fields should be massless, whereas the putative vector bosons of weak interactions, as well as the vector resonances in strong interactions all were massive. And no way was known at that time to give these particles mass.

The situation changed drastically in 1964, when [Brout and Englert, 1964], [Higgs, 1964, 1966] and [Kibble, 1967] discovered a symmetry-breaking mechanism (described in §10.3 of Bleecker's book), which allowed some components of the Yang-Mills field to acquire a mass (more precisely, in perturbation theory, allowed the pole in the propagator of the Yang-Mills field to shift away from zero), without worsening the renormalizability properties of the corresponding field theory. This led [Weinberg, 1967] and [Salam, 1968] to revive a unified model of weak and electromagnetic interactions, based on the gauge group $SU(2) \times U(1)$ and previously proposed by [Glashow, 1961] and [Salam and Ward, 1961] which is now known as the standard model and was sanctioned by the Nobel prize awards to Glashow, Salam, and Weinberg in 1979. Since about 1970 the development of gauge theories was so rapid, and the literature so vast, that it cannot be adequately covered here. The only important point I want to mention was the proof of renormalizability of the theory (important for any quantum theory which aspires to describe reality) by 't Hooft and Veltman and the discovery of topologically nontrivial solutions (instantons and monopoles) by ['t Hooft, 1973], [Polyakov, 1974], and [Belavin et al., 1975], which led to the fruitful intervention of the mathematicians and to nontrivial mathematical results (cf. Atiyah-Hitchin-Singer, 1978), and the reviews by Atiyah and Jaffe at the Helsinki Mathematical Congress, [Atiyah, 1979], [Jaffe, 1979] for further references.

To end this rather incomplete and biased history of the subject—I would like to apologize here to anyone not given due credit, or whose work has not been quoted—for a more complete and different perspective, cf. [Yang, 1979]—let me note that the idea to use fiber bundles and connections in gauge theories appeared in the 1960s ([Lubkin, 1963], [Mayer, 1965], [Hermann, 1967]), but
was not accepted till about 1973 ([Trautman, 1971, 1973], [Mayer, 1973, 1975, 1977], [Wu-Yang, 1975, 1976]). It is interesting to note that the adoption of this language was hampered by the different meaning of the word “trivial” in mathematics and in physics: whenever we asked our mathematical colleagues whether fiber bundles could help us in understanding gauge theories we were told that our bundles are “trivial”—reason enough not to pursue this subject for a while. Only after Wu and Yang showed that the Bohm-Aharonov effect, and the understanding of Dirac monopoles in electrodynamics leads to non-trivial bundles, did the notion of fiber bundle make headway into mathematical physics. It became particularly popular after the discovery of instantons, and of the results on instanton counting and the construction of instanton solutions by means of algebraic geometry methods [Atiyah-Hitchin-Singer, 1978], [Atiyah, 1979], [Atiyah, Drinfel’d, et. al., 1979], [Bourguignon, et. al., 1979], to quote just a few representative mathematical papers.

In order to understand the role gauge theories play in modern physics it is necessary to have a rudimentary understanding of how most theoretical physicists view the structure of matter at this point in time. The description is necessarily oversimplified, to fit into a mathematical framework, and no proofs exist that it is correct, but a view of this kind is currently accepted by a large fraction of the physics community. The fundamental building blocks of matter (at this point in history, since these concepts evolve with time) are the fundamental fermions: leptons and quarks. The leptons are directly observable (presently known are the electron and its neutrino, the muon and its neutrino, and the tau and its putative neutrino; in all three “flavors” or families, which behave similarly under weak and electromagnetic interactions and differ only in their masses). The quarks (which also come in three families \((u, d), (s, c), (t, b)\), standing respectively for: up, down, strange, charmed, top, bottom) are assumed to have fractional electric charges \((+2/3, -1/3\) electron charges) and are assumed to be unobservable. Each quark comes in three “colors”, and the only physically observable states are supposed to be colorless. This is described by saying that the quarks are invariant under an \(SU(3)\) “color” group, and their interactions are mediated by an eight-component gauge field—gluons—which is also unobservable, since it carries color. Observable objects are formed either of quark-antiquark pairs (e.g. the positive pion is composed of the up-quark and the down-antiquark) or of a colorless superposition of three quarks (e.g., the proton consists of two \(u\)-quarks and a \(d\)-quark, all in different color states) held together by the exchange of gluons. The leptons are capable only of weak and electromagnetic interactions. The weak interaction is thought to be mediated by the exchange of particles of the weak gauge fields: the charged \(W\)—the discovery of which was announced in January of 1983—and the yet to be confirmed neutral \(Z\), particles which are assumed to acquire masses through the Higgs mechanism. The only gauge field which remains massless is the photon \(A\), responsible for electromagnetism. In addition to interacting weakly and electromagnetically (by the exchange of these same particles) the quarks also interact strongly (by exchange of gluons). The unobservability of quarks and gluons, known as confinement, is believed to be an intrinsic feature of the unbroken gauge theory of color \(SU(3)\).
Even bolder attempts at "grand unification" are popular in physics these days. Here it is assumed that the original theory is based on a principal bundle with a larger structure group, e.g., $SU(5)$ or $SO(10)$, and that the quarks and leptons are described by a section of an associated vector bundle with five-dimensional (respectively, real ten-dimensional) fiber. Symmetry breaking then leads to the structure group $U(1) \times SU(2) \times SU(3)$ of the electroweak + color gauge model, which ultimately breaks down to the (electromagnetic $U(1)) \times (\text{color } SU(3))$ gauge bundle. Although the model is far from being a theory, it is held that it contains the germ of a future theory. Mathematically, such a theory consists of a principal bundle over spacetime manifold (suitably compactified by admitting only classical solutions which fall off at infinity to make certain "energy" integrals finite) and various associated bundles, the sections of which describe the various particles. The interactions are mediated by connections, and these in turn are determined, at least at the classical level, which can be described in usual mathematical terms, by the critical points of certain "action" functionals—four-forms integrated over the spacetime manifold, and constructed out of the curvature and its Hodge dual. This is, in a nutshell, the classical picture underlying any attempt to quantize such a theory. I will not attempt to describe quantum gauge theories, both because these are not discussed in Bleecker's book, and because the mathematical description is much more complex (cf., e.g., the reviewer's lectures, [Mayer, 1981], and the book by [Faddeev-Slavnov, 1980], for a description of quantization problems).

This brings us to the contents of the book under review. Bleecker has successfully limited his attention to the fundamental geometric aspects of classical gauge theory, and has managed to condense into 180 pages a substantial amount of information, without sacrificing the pedagogical aspects of the subject. Most proofs and calculations are carried out in detail, and thus the book should present only minor difficulties even to a graduate student in physics, who is willing to follow every word, and do every calculation. However, some previous intuition on differential forms, Lie algebras, and calculus of variations will make the reading much more enjoyable.

Chapter 0, Preliminaries, is an extremely useful summary of the necessary background in multilinear algebra, differential forms, tensors, Lie algebras, and Lie groups. The basic definitions of principal fiber bundles and connections can be found in Chapter 1. Chapter 2 is devoted to Lie-algebra valued differential forms and the concept of curvature, and gives enough details so that an inexperienced reader can follow the reasoning easily. Chapter 3 discusses particle fields, not as sections of associated vector bundles, but via the equivalent description as equivariant maps from the principal bundle to vector spaces carrying representations of the structure group. §3.2 contains a discussion of the automorphisms of the principal bundle on which a gauge theory is based and gauge transformations are defined as vertical bundle automorphisms. The reader is warned about the variations in terminology, particularly, that used in the physics literature. §3.3 deals elegantly, if succintly, with the variational principles underlying classical gauge theories, and the concept of invariant Lagrangian. The physicist-reader will find some motivations for the definitions introduced here and in the sequel lacking, but if
he is willing to provide his own motivations, will be well rewarded by a deeper understanding of the mathematics.

Chapter 4 is devoted to the principle of least action and Lagrange's equations for particle fields, making use of Hodge duality on vector bundles with connections. The author cleverly uses the metric induced on horizontal vectors by the base space metric and the metric of the vector space in which the particle fields take their values, in order to set up this duality and to introduce the notion of covariant codifferential. This leads to a particularly streamlined discussion of variational principles (maybe too streamlined for the tastes of traditionally educated theoretical physicists, who feel uncomfortable without a plethora of indices and many integrations by parts). The same remark applies to Chapter 5, which deals with the Yang-Mills current (the source term of the inhomogeneous Yang-Mills equation which is written as the covariant codifferential of the Hodge-dual of the curvature equated to the current). After some work on specific examples, the physicist will convince himself that this is indeed the particle current, and that only the total current consisting of the Yang-Mills current and the "self-current" of the connection (i.e., the term proportional to $A \wedge F$) is conserved.

Chapter 6 deals with the Dirac field describing the electrons and other fermions. The treatment contains some pedagogical innovations. It should be noted that §6.2 contains the only annoying misspelling in the book: Levi-Civita is misspelled as Levi-Cevita (this has been corrected in Chapter 8 and the Index, but survived in the Table of Contents, a surprising fact in this age of electronic spelling-checking). Chapter 7 discusses interactions between particles and gauge fields in terms of "bundle splicing" (fibered products). The formalism is illustrated by a derivation of the Dirac-Maxwell, and Dirac-Yang-Mills system (for the $SU(2)$ theory).

In preparation for general-relativistic extensions, Chapter 8 is devoted to a quick review of tensor analysis on a (pseudo)-Riemannian manifold. The unification of gauge fields and gravitation is discussed in Chapter 9. After an elegant coordinate-free preparation the author discusses nonabelian generalizations of the Kaluza-Klein theory (in which it was unsuccessfully attempted to unify electromagnetism and gravity on a five-dimensional manifold with compact fifth dimension–isomorphic to the electromagnetic gauge group $U(1)$). The derivation of the Einstein-Yang-Mills equations for the nonabelian Kaluza-Klein model proposed by the author is based on an action principle where the action density is the scalar curvature of the bundle metric. It should be noted that models of the Kaluza-Klein type, as well as other gauge theories with higher-dimensional base spaces leading to symmetry breaking via "dimensional reduction" have been much discussed in the literature in the years since the book was written (for additional references, cf., e.g., [Mayer, 1979, 1981]). The final chapter is devoted to various additional topics, such as the motion of particles in gauge potentials, Utiyama's theorem (stating that the action density of a gauge-invariant theory must be an $\text{Ad}$-invariant function of the curvature), a geometric description of the "Higgs mechanism", and a brief introduction to characteristic classes, monopoles and instantons.
The author has not set as his goal to cover all the latest development in gauge theory, particularly those which cannot yet be written in terms of rigorous mathematics (such as quantization, Feynman path integration, the Feynman-DeWitt-Faddeev-Popov trick), or those that require a deeper understanding of the index theorem (counting and construction of instantons). Those areas which are covered in the book form a useful basis for further studies. The list of references is by no means complete, and should be viewed more as a guide to further reading. The interested reader might well want to consult some of the review articles listed at the end of this review, for more complete references. The index of notations and the subject index are helpful in finding one's way through this tightly written text.

In all, Bleecker has succeeded in presenting a difficult subject succinctly and elegantly. Physicists will fault him for not providing enough motivation (mathematics texts seldom do) and mathematicians should be warned that the physics presented is sometimes oversimplified, without reference to the remaining tremendous difficulties. Nevertheless, both physicists and mathematicians will find a lot of useful material assembled in this text, which is a valuable supplement to the review literature written by physicists. The author, the series editors, and the publisher should be congratulated for producing in a relatively short time a nicely typeset text with few misprints (excepting the misspelling mentioned above). It should be hoped that the new series which this book inaugurates will continue helping bridge the gap between mathematics and the physical sciences.

I would like to thank Professors Meyer Jerison and Robert Reilly for pointing out some errors in the first draft of this review and for suggesting some improvements. Any surviving errors or inaccuracies, particularly as regards the history of the subject, are my sole responsibility.

References


BOOK REVIEWS


[Weyl, 1955] Selecta Hermann Weyl, Birkhäuser, Basel. Also, Collected papers, where the above references are reproduced.


ADDITIONAL REVIEW ARTICLES

M. Daniel and C. M. Viallet, Rev. Mod. Phys. 52 (1980), 175.

T. Eguchi, P. B. Gilkey, and A. J. Hanson, Physics Reports 66 (1980).

Additional physical insight can be gained from the many excellent articles on the subject which have appeared over the past decade in Scientific American (articles by Glashow, Nambu and others).

M. E. MAYER


Riemann surfaces, those old and venerated structures, show their smiling faces in many different connections, from the geometry of algebraic curves to the integration of nonlinear partial differential equations in mathematical physics. Even with all that is familiar, each generation finds frontiers beyond, exciting the explorers with a unique combination of explicitness and richness of technique: algebraic, analytic, and geometric. The great masters of the 19th century (Abel, Riemann,...) left a wealth of information and insight on