45.

From: GENERAL RELATIVITY AND GRAVITATION, VOL. 1
Edited by A. Held
(Plenum Publishing Corporation, 1980)

9

Fiber Bundles, Gauge Fields, and Gravitation

ANDRZEJ TRAUTMAN

1. Introduction and Motivation

1.1. Physicists were using concepts that are now part of the theory of fiber bundles before mathematicians introduced the notion of a bundle. For example, the phase space of classical mechanics and statistical physics coincides with the cotangent bundle of a configuration space. The derivation by Dirac of the formula for the strengths of magnetic poles is equivalent to the classification of circle bundles over S_2 by their Chern numbers.[1] In this respect, the situation of physicists can be likened to that of Monsieur Jourdain *qui fait de la prose sans le savoir.* There thus arises the question whether it is worth while to learn the language and use the methods of fiber bundles since so far it has been possible to do without them. It is hard to give a straightforward and convincing answer to this question; probably the only reasonable thing to say is "the future will tell." My personal opinion is that at least some concepts of fiber bundle theory will become an established part of mathematical physics because fiber bundles provide a natural and convenient framework for discussing the concepts of relativity and invariance, describing gravitation and other gauge fields, defining the notion of induced representations, and giving a geometrical interpretation to quantization and the canonical formalism of particles and fields. Fiber bundles provide a convenient language for dealing with local problems of differential geometry and field theory. They are necessary to understand and solve global, topological problems, such as those arising in connection with magnetic poles and instantons.

For a long time, Einstein searched for a "unified" geometrical theory of gravitational and electromagnetic forces. The success of his attempts, based

ANDRZEJ TRAUTMAN • Institute of Theoretical Physics, Warsaw University, ul. Hoża 69, 00-681 Warsaw, Poland.

on modifications of Riemannian geometry, was limited. Probably the best geometrical—but hardly unified—theory of this type is the one due to Kaluza and Klein. Its underlying geometry is that of a five-dimensional Riemannian space with a one-parameter group of isometries. It turns out that the Kaluza–Klein space is the total space of a circle bundle and that the electromagnetic potentials play a double role: they define a connection form on the bundle and, together with the metric of space–time, determine the five-dimensional Riemannian geometry. Gauge theories such as those based on $SU(n)$ groups have a similar geometry. If the present views on the role of gauge fields in strong and weak interactions are confirmed, then fiber bundles with connections will provide the framework for a geometrical description, according to one pattern, of all fundamental physical forces. This unification will be considerably different from Einstein's own attempts, but may be close in spirit to his program of geometrizing physics.

1.2. The notion of a fiber bundle generalizes that of a Cartesian product.(2) Two simple examples from physics and geometry will clarify the need for such a generalization.(3)

(i) In Aristotelian physics both space and time are absolute, every event being defined by an instant of time and a location in space. This is equivalent to saying that space–time E is a Cartesian product $T \times S$, where T is the time axis and S is the three-dimensional space. In Galilean physics time remains absolute, but space is relative. This can be described by saying that there is a *projection* $\pi: E \to T$, i.e., a surjective (onto) map π that associates to any event $p \in E$ the corresponding instant of time $t = \pi(p) \in T$. The set (line) T is called the *base space* and the set $\pi^{-1}(t)$ of all events simultaneous with p is called the *fiber* over t. Each fiber is isomorphic to the Euclidean three-dimensional space $\mathbf{R}^3$, which is therefore called the *typical fiber*. The *total space* E of this fiber bundle may be trivialized, i.e., represented as the Cartesian product $T \times \mathbf{R}^3$. Any such trivialization (map) $h: E \to T \times \mathbf{R}^3$ is of the form $h(p) = (\pi(p), \mathbf{r}(p))$, where $\mathbf{r}(p) = (x(p), y(p), z(p))$ are the space coordinates of the event p relative to an inertial observer. One can say that Galilean space–time E is the total space of a fiber bundle which is *trivial*, i.e., isomorphic to the product bundle $T \times \mathbf{R}^3$, without there being a *natural* isomorphism between these bundles.

(ii) Consider the two-dimensional sphere S_2 with a preferred orientation. Define a "dyad" as a pair of unit orthogonal vectors tangent to S_2 at a point. Let P be the set of all dyads whose orientation agrees with that of S_2. One can make P into the total space of a bundle in such a way that $\pi: P \to S_2$ is the map sending a dyad into the point at which its vectors are attached to S_2. If $e = (e_1, e_2)$ is a dyad at $x \in S_2$, then so is the pair (e_1', e_2'), where

$$e_1' = e_1 \cos\varphi + e_2 \sin\varphi, \qquad e_2' = -e_1 \sin\varphi + e_2 \cos\varphi \tag{1.1}$$

and all dyads at x may be obtained in this manner from (e_1, e_2). Therefore, $SO(2)$ is the typical fiber of the bundle $\pi: P \to S_2$. Equation (1.1) defines an *action* of the (*structure*) *group* $SO(2)$ on P. The bundle $\pi: P \to S_2$ is a simple example of a *principal bundle*. Moreover, this bundle is nontrivial in the following sense: there is no diffeomorphism $k: S_2 \times SO(2) \to P$ such that $\pi \circ k(x, a) = x$. Indeed, if such a k existed, then $s: S_2 \to P$, defined by $s(x) = k(x, a_0)$, would determine a smooth field of unit vectors on S_2. By the "no combing of S_{2n}" theorem of Brouwer, such a field s does not exist. In general, if $\pi: E \to M$ is a bundle and N is an open subset of M, then a smooth map $s: N \to P$, such that $\pi \circ s = \mathrm{id}_N$, is called a (local) *section* of π. If $N = M$ then s is a *global section*. For a principal bundle, the existence of a global section is equivalent to its triviality. Incidentally, the bundle of dyads occurs in the description of a magnetic pole of unit strength. (The system of physical units used here is such that the charge of the electron is equal to the fine-structure constant.) The nontrivial nature of the bundle $\pi: P \to S_2$, shows up in the occurrence of a "string singularity" in the expression for the vector potential of the magnetic pole.[4]

1.3. The last remark leads to what is probably the most important domain of applications of fiber bundles in theoretical physics: infinitesimal connections on principal bundles provide good geometrical models of classical gauge fields. This has been known among mathematicians and physicists for some time but, for the sake of completeness, let us recall some of the arguments in favor of this view. In a notation that is standard in physics, one can consider the analogies between electromagnetism and gravitation:

Electromagnetism:	Gravitation:
$A'_\mu = A_\mu + \partial_\mu \chi$	$\Gamma'^{\mu}_{\nu\rho} = \Gamma^{\alpha}_{\beta\gamma} \dfrac{\partial x'^\mu}{\partial x^\alpha} \dfrac{\partial x^\beta}{\partial x'^\nu} \dfrac{\partial x^\gamma}{\partial x'^\rho} + \dfrac{\partial x'^\mu}{\partial x^\alpha} \dfrac{\partial^2 x^\alpha}{\partial x'^\nu \partial x'^\rho}$
$\partial_\mu - iA_\mu$	∇_μ
$F_{\mu\nu}$	$R^\alpha{}_{\beta\mu\nu}$
$F_{[\mu\nu,\rho]} = 0$	$R^\alpha{}_{\beta[\mu\nu;\rho]} = 0$

The issues raised in the discussion on the significance of the electromagnetic potentials become clear when electromagnetism is interpreted as an (infinitesimal) connection in the space of phases. Namely, the experiments proposed by Aharonov and Bohm[44] have a very simple analog in elementary geometry: the surface of a cone is locally flat, but a vector undergoing parallel transport along a loop enclosing the vertex does not return to its original position. Similarly the phase of a wave function of a charged particle undergoes parallel transport determined by the potential. The region with the magnetic field is analogous to the vertex of the cone. Electromagnetic potentials should not be slighted, but considered for what they are: the coefficients of a connection.

A heuristic approach to the notion of a connection on a principal bundle shows how this concept is related to the physicist's view of gauge potentials: Let $\pi: P \to M$ be a principal bundle with structure group G. The result of action of $a \in G$ on $p \in P$ is another point $pa \in P$, lying in the same fiber as p, $\pi(pa) = \pi(p)$. A local section $s: N \to P$ defines a diffeomorphism $k: N \times G \to \pi^{-1}(N)$ by $k(x, a) = s(x)a = p$. The section s being fixed for the moment, we may identify $s(x)$ with (x, ε) and $s(x)a$ with $(x, a) = (x, \varepsilon)a$, where ε is the unit element of G. An infinitesimal connection on P defines parallel displacement of elements of P. If $dx = (dx^\mu)$ is a small displacement at $x = \pi(p) \in N$, then the parallel transport of (x, ε) along dx results in $(x + dx, \varepsilon - A)$, where $A = A_\mu\, dx^\mu$ is a 1-form on N, with values in the Lie algebra G' of G. Parallel transport should commute with the action of G: (x, a) displaced along dx becomes $(x + dx, a - Aa)$. If $s': N' \to P$ is another section, then there is a map $U: N \cap N' \to G$ such that

$$s'(x) = s(x)U(x) \tag{1.2}$$

for $x \in N \cap N'$. The section s' leads to the diffeomorphism $k': N' \times G \to \pi^{-1}(N')$, $k'(x, a) = s'(x)a = s(x)U(x)a$, and

$$k'(x, a) = k(x, Ua)$$

$$k'(x + dx, a) = k(x + dx, (U + dU)a)$$

Relative to k', parallel transport is described by a 1-form $A' = A'_\mu\, dx^\mu$. By parallel transport, the point $k'(x, \varepsilon)$ becomes $k'(x + dx, \varepsilon - A')$, which is the same as $k(x + dx, (U + dU)(\varepsilon - A'))$. On the other hand, $k'(x, \varepsilon) = k(x, U)$ is parallel to $k(x + dx, U - AU)$. Since parallel displacement in P should not depend on the choice of section (gauge), $(U + dU)(\varepsilon - A') = U - AU$. This leads to the transformation law

$$A' = U^{-1}(dU + AU) \tag{1.3}$$

of the potential under gauge transformations of the second kind. It follows from (1.3) that the G'-valued 1-form

$$\omega = a^{-1}(da + Aa) \tag{1.4}$$

is independent of the section. The form ω has a simple geometric interpretation: $\varepsilon + \omega$ is the element of G that moves the point (x, a) into the point $(x, a)(\varepsilon + \omega) = (x, a + da + Aa)$ parallel to $(x + dx, a + da)$. The section-independent 1-form ω on P is called the connection form; it is the gauge-independent counterpart of the potential A. Relation (1.3) contains, as special cases, the transformation laws of the coefficients of a linear connection (Christoffel symbols, Ricci rotation coefficients), of the electromagnetic potentials, and of non-Abelian gauge potentials of the Yang–Mills type. The advantage of the connection form ω, defined on P, over the potential A, defined on $N \subset M$, results from the following considerations: the connection form ω is defined independently of any section, whereas A refers to a (local) section of the bundle. As a consequence, for a nontrivial bundle, the potentials are defined only locally, whereas the connection form ω is defined globally, all over P.

1.4. An interesting application of the bundle approach to gauge fields is the construction of Riemannian geometries of the Kaluza–Klein type.[3] If there is a connection form ω on P, $g = g_{\mu\nu}\, dx^{\mu}\, dx^{\nu}$ is a metric tensor on M and h is a bi-invariant metric on G, then one can define a metric tensor γ on P by the formula

$$\gamma(u, v) = g\big(T\pi(u), T\pi(v)\big) + \text{const } h\big(\omega(u), \omega(v)\big)$$

where u and v are vectors tangent to P, and $T\pi\colon TP \to TM$ is the projection of such vectors on M, induced by π. The metric γ is invariant under the action of G on P. For $G = SO(2)$ it coincides with the metric considered in five-dimensional, "unified" theories of gravitation and electromagnetism.

1.5. Relativistic theories of gravitation—such as Einstein's theory of general relativity—may also be considered as gauge theories. The bundle P consists in this case of orthonormal linear frames (tetrads, *Vierbeine*) of the space–time manifold M and G is the Lorentz group. Alternatively, one can take P to be the bundle of orthonormal affine frames, in which case G is the inhomogeneous Lorentz group. There are, however, important differences between Einstein's theory and gauge theories such as electrodynamics or the Yang–Mills theory. First of all, the bundle of frames is *soldered* to the base M whereas in other gauge theories the bundle is rather loosely connected to M (Section 4.1).

The soldering results in the appearance, in theories of gravitation, of *torsion*, in addition to *curvature*, which occurs in any gauge theory. (Torsion is zero in Riemannian geometry, but being zero is different from not existing at all.) Moreover, the form of Einstein's equations of gravitation is different from the "generic" form of the field equations assumed in gauge theories. The latter are derived from Lagrangians quadratic in curvature, whereas the former are based on a linear Lagrangian. The possibility of constructing such a linear Lagrangian is also related to the existence of the soldering form on P.

2. Fiber Bundles and Infinitesimal Connections

2.1. In physics, we usually need spaces on which differential calculus can be developed. For this reason, we restrict our considerations to the differentiable case. The reader should consult the books and articles listed at the end of the paper for precise definitions and properties of bundles with connections.[5–8]

2.2. A smooth principal bundle includes, in its definition, the following list of differentiable manifolds and smooth maps:

a total (bundle) space P

a Lie group G

a base space M

a projection $\pi: P \to M$

a map $\delta: P \times G \to P$ defining the action of G on P to the right; if $a, b \in G$ and $\varepsilon \in G$ is the unit element, then

$$\delta(a) \circ \delta(b) = \delta(ba) \quad \text{and} \quad \delta(\varepsilon) = \mathrm{id}$$

where $\delta(a)p = \delta(p, a)$, $p \in P$; moreover

$$\pi \circ \delta(a) = \pi$$

One often writes pa instead of $\delta(a)p$.

2.3. A connection is given by a 1-form ω on P, with values in G', the Lie algebra of G. For any $v \in G'$, $\delta(\exp tv)$ is the one-parameter group of transformations of P generated by v; if $\Delta(v)$ is the vector field on P induced by $\delta(\exp tv)$, then

$$\omega[\Delta(v)] = v$$

and

$$\delta(a)^*\omega = \mathrm{ad}\,(a^{-1})\omega \tag{2.1}$$

where $\delta(a)^*\omega$ is the pull-back of ω by $\delta(a)$, i.e., $\delta(a)^*\omega(u) = \omega[T\delta(a)u]$ for any $u \in TP$. $\mathrm{ad}\,(a)$ is the automorphism of G' associated to $a \in G$ by the adjoint representation of G in its Lie algebra.

At any point $p \in P$ one defines the horizontal subspace by

$$\mathrm{hor}_p\, P = \{u \in T_pP: \omega(u) = 0\}$$

There is a direct sum decomposition

$$T_pP = \mathrm{hor}_p\, P \oplus \mathrm{ver}_p\, P \tag{2.2}$$

where the vertical space $\mathrm{ver}_p\, P$ coincides with the kernel of the map $T_p\pi$, tangent to π at p. Therefore, on a bundle with connection, a vector $u \in T_pP$ admits a decomposition

$$u = \mathrm{hor}\, u + \mathrm{ver}\, u$$

corresponding to (2.2).

2.4. Let $\rho: G \to GL(V)$ be a representation of G in a vector space V. A k-form α on P with values in V is said to be of type ρ if it is equivariant under the action of G,

$$\delta(a)^*\alpha = \rho(a^{-1})\alpha$$

If α is such a form, then so is the form hor α defined by

$$\text{hor}\,\alpha(u_1, \ldots, u_k) = \alpha(\text{hor}\,u_1, \ldots, \text{hor}\,u_k)$$

where

$$u_1, \ldots, u_k \in T_pP$$

If α is a k-form of type ρ, then its covariant exterior derivative

$$D\alpha = \text{hor}\,d\alpha$$

is a $(k+1)$-form of the same type. For example,

$$\Omega = D\omega$$

is a 2-form of type ad, called the curvature form. Explicitly,

$$\Omega = d\omega + \tfrac{1}{2}[\omega, \omega] \tag{2.3}$$

where the commutator of two G'-valued forms is defined as follows: Let (e_i) be a frame (basis) in G', and let $\alpha = \alpha^i e_i$ and $\beta = \beta^i e_i$ be two such forms; then

$$[\alpha, \beta] = \alpha^i \wedge \beta^j [e_i, e_j] = \alpha^i \wedge \beta^j c^k{}_{ij} e_k$$

where $(c^k{}_{ij})$ are the structure constants of G with respect to (e_i). There always holds the *Bianchi identity*

$$D\Omega = 0 \tag{2.4}$$

If α is a horizontal k-form of type ρ,

$$\text{hor}\,\alpha = \alpha$$

then $D\alpha$ may be evaluated from the formula

$$D\alpha^a = d\alpha^a + \rho^a_{bi}\omega^i \wedge \alpha^b \tag{2.5}$$

where

$$\alpha = \alpha^a e_a, \qquad D\alpha = (D\alpha^a)e_a, \qquad \omega = \omega^i e_i$$

(e_a) is a frame in V, and ρ^a_{bi} is the ath component of the vector

$$\frac{d}{dt}\rho(\exp te_i)e_b\bigg|_{t=0}$$

2.5. A ρ-invariant metric on V is a bilinear symmetric map

$$h: V \times V \to \mathbf{R}$$

such that

$$h(\rho(a)u, \rho(a)v) = h(u, v) \quad \text{for any } a \in G \text{ and } u, v \in V$$

If G is connected, then h is ρ invariant iff

$$h_{ac}\rho^{c}_{\ bi} + h_{cb}\rho^{c}_{\ ai} = 0$$

where $h_{ab} = h(e_a, e_b)$. For example, the Killing metric on G' is ad invariant.

2.6. *Associated bundles.* Let $\rho: G \times F \to F$ denote an action of G in a manifold F to the left, $\rho(a) \circ \rho(b) = \rho(ab)$, $\rho(\varepsilon) = \mathrm{id}$ where $\rho(a)q = \rho(a, q)$. If $\pi: P \to M$ is a principal G-bundle, then one defines an action of G in $P \times F$ by

$$(p, q)a = (\delta(p, a), \rho(a^{-1}, q))$$

Let

$$\kappa: P \times F \to E = (P \times F)/G$$

be the canonical map on the quotient of $P \times F$ by G. The set E has a natural structure of a fiber bundle with projection $\pi_E: E \to M$ defined by $\pi_E \circ \kappa(p, q) = \pi(p)$. The manifold F is the typical fiber of the bundle E associated to P by ρ. For example, if V is a vector space and $\rho: G \to GL(V)$ is a representation of G in V, then each fiber of E has a natural structure of a vector space and E is said to be a *vector bundle*.

Let $\varphi: P \to F$ be a mapping equivariant with respect to the action of G in P and F, i.e., such that for any $a \in G$

$$\varphi \circ \delta(a) = \rho(a^{-1}) \circ \varphi \tag{2.6}$$

Define the section $\tilde{\varphi}$ of $\pi_E: E \to M$ by

$$\tilde{\varphi}(x) = \kappa(p, \varphi(p)) \tag{2.7}$$

where $p \in \pi^{-1}(x)$; the right side of (2.7) does not depend on the choice of p inside the fiber over $x \in M$. Conversely, given a section $\tilde{\varphi}: M \to E$ of π_E, one can define an equivariant mapping $\varphi: P \to F$ by

$$\varphi(p) = \kappa_p^{-1} \circ \tilde{\varphi} \circ \pi(p)$$

where

$$\kappa_p: F \to \pi^{-1}(\pi(p))$$

is the diffeomorphism defined by $\kappa_p(q) = \kappa(p, q)$. Therefore, there is a natural, one-to-one correspondence between equivariant mappings from P to F and sections of the associated bundle with fiber F.

2.7. *Example.* Let $P = LM$ be the bundle of linear frames of an n-dimensional manifold M. Its structure group G is $GL(n, \mathbf{R})$ and the action is given by $(e, a) \mapsto \delta(e, a) = ea$, where $e = (e_\mu)$, $a = (a^\mu{}_\nu)$, $ea = (e_\mu a^\mu{}_\nu)$, and $\mu, \nu = 1, \ldots, n$. For any linear representation (homomorphism) $\rho: GL(n, \mathbf{R}) \to GL(V)$ one constructs the associated vector bundle $E = \rho(M)$ of quantities of type ρ over M. A section $\tilde{\varphi}: M \to E$ of this bundle is a *field* (of quantities) of type ρ and $\varphi(e) \in V$ gives the *components* of the field with respect to the frame e at $x = \pi(e)$. A section $s: M \to P$ of P is a *field of frames* and $\Phi = \varphi \circ s: M \to V$ is the expression of the field by its components relative to s. If $U: M \to G$ then $s' = sU$ is another field of frames and, according to (2.6), $\Phi' = \varphi \circ s'$ is related to Φ by the *transformation law* of components of a field of type ρ,

$$\Phi(x) = \rho\big(U(x)\big)\Phi'(x)$$

2.8. Definition of principal bundles by transition functions is close to the physicist's way of thinking in terms of local coordinates, gauge transformations, etc. Since any $x \in M$ has a neighborhood N such that $\pi^{-1}(N)$ is isomorphic to $N \times G$, it is possible to find an open covering (N_α) of M and diffeomorphisms $k_\alpha: N_\alpha \times G \to \pi^{-1}(N_\alpha)$ such that $\pi \circ k_\alpha(x, a) = x$ and $k_\alpha(x, a)b = k_\alpha(x, ab)$. If $x \in N_\alpha \cap N_\beta$, then there is an element $a_{\alpha\beta}(x)$ of G such that $k_\beta(x, \varepsilon) = k_\alpha\big(x, a_{\alpha\beta}(x)\big)$ and

$$a_{\alpha\beta}: N_\alpha \cap N_\beta \to G \tag{2.8}$$

is a mapping. Moreover

$$\text{if} \quad x \in N_\alpha \cap N_\beta \cap N_\gamma \quad \text{then} \quad a_{\alpha\gamma}(x) = a_{\alpha\beta}(x)a_{\beta\gamma}(x) \tag{2.9}$$

Clearly, there is a great deal of arbitrariness in the choice of the diffeomorphisms k_α: given a family (U_α) of mappings $U_\alpha: N_\alpha \to G$, one can define the "transformed" diffeomorphisms $k'_\alpha(x, a) = k_\alpha\big(x, aU_\alpha(x)\big)$ and the "transformed" transition functions

$$a'_{\alpha\beta}(x) = U_\alpha(x)^{-1}a_{\alpha\beta}(x)U_\beta(x) \quad \text{for} \quad x \in N_\alpha \cap N_\beta \tag{2.10}$$

Conversely, given an open covering (N_α) of a manifold M and a family $(a_{\alpha\beta})$ of transition functions (2.8) subject to (2.9), there is a principal bundle $\pi: P \to M$ with group G, and a family (s_α) of local sections, $s_\alpha: N_\alpha \to P$ such that $s_\beta(x) = s_\alpha(x)a_{\alpha\beta}(x)$ for $x \in N_\alpha \cap N_\beta$. The bundle is determined by the transition functions uniquely up to isomorphisms; two families of transition functions related by (2.10) lead to isomorphic bundles.

2.9. There is a general method of constructing principal bundles from Lie groups. Let us first recall that, if H is a closed Lie subgroup of a Lie group G, then the set G/H of left cosets modulo H has the structure of a manifold. The group G acts in G/H to the left by $a(bH) = abH$, where bH is the coset containing $b \in G$. This action of G in G/H is transitive and any manifold F *homogeneous* under an action ρ of G may be obtained in this manner by taking for H the *stability* (*isotropy*) group of one of its points, say o

$$H = \{a \in G : \rho(a)o = o\}$$

Moreover, $G \to G/H$ is a principal H-bundle. To describe the more general construction, consider two closed Lie subgroups K and H of G such that K is a normal subgroup of H. The quotient H/K is then a Lie group and there is a principal H/K-bundle

$$\pi : G/K \to G/H$$

defined by $\pi(aK) = aH$, $\delta(aK, bK) = abK$ where $a \in G$, $b \in H$.

3. Gauge and Higgs Fields

3.1. To construct a gauge theory[9] it is necessary to specify (i) a gauge group G; (ii) the type of particles that are coupled to the gauge field; this is done by choosing a representation ρ of G in V; (iii) the form of the field equations. One then considers principal G-bundles over spacetime M and connections defined over these bundles. Let ω be a connection form on a G-bundle $\pi : P \to M$. The potential corresponding to a local section $s : N \to P$ is

$$A = s^* \omega$$

If (x^μ) is a system of local coordinates in $N \subset M$, then

$$A = A^i_\mu \, dx^\mu \, e_i$$

Similarly, the *field strengths* relative to the local section s are

$$F = s^* \Omega = \tfrac{1}{2} F^i{}_{\mu\nu} \, dx^\mu \wedge dx^\nu \, e_i$$

By virtue of (2.3) one has

$$F^i_{\mu\nu} = A^i_{\nu,\mu} - A^i_{\mu,\nu} + c^i_{jk} A^j_\mu A^k_\nu$$

A change of the local section implies a change of the potential and of the field strengths. The new potential A' is given by equation (1.3) and the corresponding formula for the field strengths is

$$F' = U^{-1} F U$$

3.2. A particle of type ρ, interacting with the gauge field, is described by a map $\varphi : P \to V$ subject to (2.6). The pull-back of φ by a section s is the physicists' *Higgs field* or wave function

$$\Phi = \varphi \circ s : N \to V$$

It is convenient to refer to φ itself as the Higgs field. According to what was said in Section 2.6, φ corresponds to a section $\tilde{\varphi}$ of the bundle associated to P by ρ; sometimes one prefers to work with $\tilde{\varphi}$ instead of φ. According to (2.5), one has the following explicit formula:

$$\nabla_\mu \Phi^a = \Phi^a{}_{,\mu} + \rho^a_{bi} A^i_\mu \Phi^b$$

for the pull-back $D\Phi = s^* D\varphi = \nabla_\mu \Phi^a \, dx^\mu \, e_a$ of the covariant derivative of φ. A standard Higgs fields corresponds to $V = G'$ and $\rho = \mathrm{ad}$, so that $\rho^i_{jk} = c^i_{jk}$.

3.3. Assume now that the base M is an oriented Riemannian space with metric g; one can then form the dual $*\,\alpha$ of any horizontal form α on P. The sum of degrees of the forms α and $*\,\alpha$ is equal to the dimension of M. Let (e_i) be a frame in G' and let k be an ad-invariant nonsingular metric on G'. The form

$$k_{ij} * \Omega^i \wedge \Omega^j \tag{3.1}$$

where $k_{ij} = k(e_i, e_j)$ and $\Omega = \Omega^i e_i$ is invariant under the action of G and may be taken as the Lagrange density of a variational principle for a pure gauge field. If $\delta\omega$ is a variation of the connection form, then

$$\delta\Omega = D\delta\omega \tag{3.2}$$

and the resulting field equation is

$$D * \Omega = 0 \tag{3.3}$$

The interaction between the gauge and the Higgs field φ is taken into account by supplementing (3.1) with a term of the form

$$h_{ab} * D\varphi^a \wedge D\varphi^b + U(|\varphi|)\eta \tag{3.4}$$

where $|\varphi|^2 = h_{ab}\varphi^a\varphi^b$, h is a ρ-invariant metric on V, and η is a horizontal volume element. One says that the coupling between φ and ω is minimal since the interaction term (3.4) contains ω only through $D\varphi$.

3.4. *Spontaneous Symmetry Breaking.* Consider a principal G-bundle P over M and a Higgs field whose range is an orbit W of G in V, i.e.,

$$\varphi: P \to W \subset V$$

and W is such that for any pair of points, w_0, $w \in W$ there is an $a \in G$ so that $w = \rho(a)w_0$. Let H be the isotropy group of w_0,

$$H = \{a \in G : \rho(a)w_0 = w_0\}$$

Then

$$Q = \{p \in P : \varphi(p) = w_0\}$$

is a subbundle of P over the same base M; its structure group is H.[10]

Conversely, given a reduction Q of the bundle P to a subgroup H of its structure group G, one defines a Higgs field $\varphi: P \to W = G/H$ by putting $\varphi(p) = H \in W$ for $p \in Q$.

A connection form ω on P, restricted to Q, defines an H-connection on Q iff it is H'-valued, i.e., iff

$$D\varphi = 0 \tag{3.5}$$

This condition is usually taken as part of the definition of the ground state in a gauge theory.

As an example of spontaneous symmetry breaking, consider the 't Hooft–Polyakov (hedgehog) solution[11–13] of the Yang–Mills equations with a Higgs field of type ad. For any fixed t and $r > 0$, the base may be identified with S_2. The $SO(3)$ bundle P over S_2 is trivial, $P = S_2 \times SO(3)$, and the (normalized) Higgs field $\varphi: P \to S_2 \subset R^3$ is $\varphi(\hat{\mathbf{r}}, a) = a^{-1}\hat{\mathbf{r}}$ for any $\hat{\mathbf{r}} \in S_2$ and $a \in SO(3)$. The north pole $\hat{\mathbf{r}}_0 = (0, 0, 1) \in S_2$ is unchanged by rotations around the z axis, thus $H = SO(2)$ and

$$Q = \{(\hat{\mathbf{r}}, a) \in P: a^{-1}\hat{\mathbf{r}} = \hat{\mathbf{r}}_0\}$$

may be identified with $SO(3)$ by $Q \ni (\hat{\mathbf{r}}, a) \mapsto a \in SO(3)$. Clearly, $SO(3) \to S_2$ is nontrivial and carries an $SO(2)$ connection corresponding to a magnetic pole with $n = 2$. It should be noted that condition (3.5) is not satisfied by the hedgehog solution. The $SO(2)$ connection is obtained here by projecting the $SO(3)$ connection on the $\hat{\mathbf{r}}_0$ direction, rather than by restriction. We see from this simple example that by a spontaneous breaking of symmetry it is possible to obtain a nontrivial bundle Q even though P is trivial.

3.5. Topological invariants for gauge fields may be obtained by the Chern–Weil construction.[14] Let

$$f: G' \times G' \times \cdots \times G' \to \mathbf{R}$$

be a k-linear symmetric map, invariant under the adjoint action of G in G'. If $\alpha_1, \ldots, \alpha_k$ are G'-valued forms on P, $\alpha_i = \alpha_i^{j(i)} e_{j(i)}$, then

$$f(\alpha_1, \ldots, \alpha_k) = \alpha_1^{j(1)} \wedge \alpha_2^{j(2)} \wedge \cdots \wedge \alpha_k^{j(k)} f(e_{j(1)}, e_{j(2)}, \ldots, e_{j(k)})$$

is an invariant, $\mathbf{R}$-valued form on P. We write

$$f(\alpha) = f(\alpha, \alpha, \ldots, \alpha),$$

$$f(\alpha, \beta) = f(\alpha, \beta, \ldots, \beta)$$

Any two G connections on the principal bundle $P \to M$, say ω and ω_1 can be smoothly joined by a one-parameter family of connections (ω_t),

$$\omega_t = \omega + t\alpha, \qquad 0 \le t \le 1$$

where

$$\alpha = \omega_1 - \omega$$

is a horizontal form of type ad. If Ω_t is the curvature form of ω_t, then

$$\frac{d}{dt} f(\Omega_t) = k\, df(\alpha, \Omega_t) \tag{3.6}$$

By invariance of f, the form $f(\Omega_t)$ projects to a form $f(F_t)$ defined globally on M and the same is true of $f(\alpha, \Omega_t)$. Integrating both sides of the projection of (3.6) over a cycle c in M one obtains

$$\int_c f(F) = \int_c f(F_1) \tag{3.7}$$

This is commonly interpreted to mean that the integral (3.7) is a "topological invariant" although the argument presented here shows only that the integral is unchanged by smooth deformations of the connection. Moreover, the form $f(\Omega)$ on P is exact,

$$f(\Omega) = kd \int_0^1 f\big(\omega, t\Omega + \tfrac{1}{2}(t^2 - t)[\omega, \omega]\big)\, dt$$

(The last relation does not project globally on M because ω is not horizontal.)

For example, if k is an invariant metric, one has the Pontryagin invariant associated with

$$k_{ij}\Omega^i \wedge \Omega^j = d(k_{ij}\omega^i \wedge \Omega^j + \tfrac{1}{3}c_{ijk}\omega^i \wedge \omega^j \wedge \omega^k)$$

where

$$\begin{aligned} c_{ijk} &= k_{il}c^l_{\ jk} \\ &= c_{[ijk]} \qquad \text{by invariance of } k \end{aligned}$$

3.6. Natural connections on Stiefel bundles provide an important class of pure (sourceless) gauge fields.(15) Let $G(n)$ be one of the three groups $SO(n)$, $U(n)$, or $Sp(n)$. The general construction of Section 2.9 leads to the Stiefel $G(n)$ bundle over a Grassmannian manifold,

$$G(m+n)/G(m) \to G(m+n)/[G(m) \times G(n)]$$

The $G(n)'$-valued part of the canonical form on $G(m+n)$ projects to a connection on $G(m+n)/G(m)$ which is sourceless and universal for $G(n)$ connections. The cases $U(2)/U(1) \to U(2)/[U(1) \times U(1)]$ and $Sp(2)/Sp(1) \to Sp(2)/[Sp(1) \times Sp(1)]$ correspond to the magnetic pole and the instanton of lowest order, respectively.(4)

4. Gravitation

4.1. Gravitation is different from other gauge theories in several aspects. The origin of these differences may be traced to the soldering of the bundle of linear frames LM to the base manifold M. For an n-dimensional manifold M, the soldering form[45] $\theta\colon TLM \to \mathbf{R}^n$ is defined as follows: if $e = (e_\mu) \in LM$ and $u \in T_e LM$ then $\theta^\mu(u)$ is the μth component, with respect to e, of the vector $T_e\pi(u)$, obtained by projecting u on the base. Clearly, $\theta = (\theta^\mu)$, $\mu = 1, \ldots, n$, is a 1-form of type id

$$\delta(a)^*\theta = a^{-1}\theta$$

If $s\colon N \to LM$ is a local section of $LM \to M$, i.e., a field of frames on $N \subset M$, then

$$s^*\theta^\mu = s^\mu$$

where $s^\mu(p)$ is the μth element of the frame dual to $s(p)$, $p \in N$,

$$\langle s_\nu(p), s^\mu(p)\rangle = \delta^\mu_\nu$$

For example, if (x^μ) is a system of local coordinates in N and s is the field of natural (coordinate, holonomic) frames associated to (x^μ), then $s^\mu = dx^\mu$. If $\omega = (\omega^\mu{}_\nu)$ is the form of a *linear connection* (=connection on LM), then the covariant exterior derivative of θ is the 2-form Θ of torsion,

$$\Theta^\mu = d\theta^\mu + \omega^\mu{}_\nu \wedge \theta^\nu \tag{4.1}$$

A *metric tensor* g on M defines the map

$$(g_{\mu\nu})\colon LM \to \mathscr{L}^2_s(\mathbf{R}^n, \mathbf{R})$$

given by

$$g_{\mu\nu}(e) = g(e_\mu, e_\nu)$$

4.2. If M is four dimensional, as it will be assumed from now on, one defines

$$\eta_{\mu\nu\rho\sigma}\colon LM \to \mathbf{R}$$

by

$$\eta_{0123}(e) = |\det g_{\mu\nu}(e)|^{1/2}$$

$$\eta_{\mu\nu\rho\sigma} = \eta_{[\mu\nu\rho\sigma]}$$

and also

$$\eta_{\mu\nu\rho} = \theta^\sigma \eta_{\mu\nu\rho\sigma}, \qquad \eta_{\mu\nu} = \tfrac{1}{2}\theta^\rho \wedge \eta_{\mu\nu\rho}, \qquad \eta_\mu = \tfrac{1}{3}\theta^\nu \wedge \eta_{\mu\nu}, \qquad \eta = \tfrac{1}{4}\theta^\mu \wedge \eta_\mu$$

In order to appreciate the differences between gravitation, understood as a theory based on a linear connection and a metric tensor, and a gauge theory on a principal bundle without soldering, consider some of the invariant forms that may be constructed in each of the following four cases.

(i) Suppose one has a gauge theory on a bundle $P \to M$ with structure group $G \subset GL(4, \mathbf{R})$ based on a connection form $\omega = (\omega^\mu{}_\nu)$, with no metric and no soldering. From the curvature form

$$\Omega^{\mu}{}_{\nu} = d\omega^{\mu}{}_{\nu} + \omega^{\mu}{}_{\rho} \wedge \omega^{\rho}{}_{\nu} \tag{4.2}$$

one constructs two closed forms of the type considered in Section 3.5,

$$\Omega^{\mu}{}_{\mu}, \qquad \Omega^{\mu}{}_{\nu} \wedge \Omega^{\nu}{}_{\mu}$$

Upon integration on cycles they both lead to quantities invariant under smooth deformations.

(ii) If one adds a metric g on M, then one can construct the dual $*\,\Omega^{\mu}{}_{\nu}$ and the conformally invariant Lagrangian density

$$*\,\Omega^{\mu}{}_{\nu} \wedge \Omega^{\nu}{}_{\mu} \tag{4.3}$$

(iii) If one is given a connection form ω on LM then in addition to curvature (4.2), one constructs torsion (4.1). There are no more invariant forms than in case (i).

(iv) Given ω on LM and g on M, in addition to the forms and invariants occurring in (i)–(iii), one can construct the η's and also the following:

$$\tfrac{1}{2}\eta_{\mu\nu} \wedge \Omega^{\mu\nu}, \qquad \text{the Einstein–Cartan Lagrangian} \tag{4.4}$$

$$\eta_{\mu\nu\rho\sigma}\Omega^{\mu\nu} \wedge \Omega^{\rho\sigma}, \qquad \text{the Euler form} \tag{4.5}$$

$$g_{\mu\nu} * \Theta^{\mu} \wedge \Theta^{\nu}, \qquad \text{the square of torsion} \tag{4.6}$$

$$g^{\mu\rho}g^{\nu\sigma} * Dg_{\mu\nu} \wedge Dg_{\rho\sigma}, \qquad \text{etc.} \tag{4.7}$$

It is clear that gravitation, understood as a theory based on g and ω defined over LM (last case) is richer than other gauge theories. The metric tensor may be looked upon as a Higgs field which breaks the symmetry from $GL(4, \mathbf{R})$ down to the Lorentz group. As noted by Nambu[16] equation (3.5) reduces in this case to the usual compatibility condition between a linear connection and a metric, assumed in both the Einstein and the Einstein–Cartan theories.[17] Moreover, the soldering form θ is also a kind of a Higgs field; it differs from other Higgs fields by being a 1-form (rather than a 0-form) and by being uniquely determined by M alone. If one takes these observations seriously, one may be led to consider a theory of gravitation based on a Lagrangian that is a sum of terms proportional to (4.3), (4.6), and (4.7).

4.3. Consider a manifold M with a metric g and a connection ω which need not be symmetric or even metric. By supplementing the Einstein–Cartan Lagrangian (4.4) with a 4-form representing matter and varying their sum with respect to g and ω, one is led to the system of equations[18,19]

$$\tfrac{1}{2}\eta_{\mu\nu\rho} \wedge \Omega^{\nu\rho} = -8\pi t_{\mu} \tag{4.8}$$

$$D\eta_{\mu}{}^{\nu} = 8\pi s_{\mu}{}^{\nu} \tag{4.9}$$

If the source 3-forms t_μ and $s_\mu{}^\nu$ do not depend on ω, then equations (4.8) and (4.9) are invariant under *projective transformations* of the connection,

$$\omega^\mu{}_\nu \mapsto \omega^\mu{}_\nu + \delta^\mu{}_\nu \lambda \tag{4.10}$$

In this case, among the connections related by (4.10) there is one which is *metric*

$$Dg_{\mu\nu} = 0 \tag{4.11}$$

if, and only if, $s_{\mu\nu}$ is skew,

$$s_{\mu\nu} + s_{\nu\mu} = 0 \tag{4.12}$$

A theory of space, time, and gravitation based on equations (4.8), (4.9), (4.11), and (4.12) is called the *Einstein–Cartan theory*. Similarly as in Einstein's theory, the sources may be either derived from an action principle or postulated on the basis of phenomenological or statistical considerations.

4.4. Writing

$$\Omega^\mu{}_\nu = \tfrac{1}{2} R^\mu{}_{\nu\rho\sigma}\theta^\rho \wedge \theta^\sigma, \qquad \Theta^\mu = \tfrac{1}{2} Q^\mu{}_{\rho\sigma}\theta^\rho \wedge \theta^\sigma$$

$$R_{\mu\nu} = R^\rho{}_{\mu\rho\nu}, \qquad R = g^{\mu\nu}R_{\mu\nu}$$

$$t_\mu = \eta_\nu t^\nu{}_\mu, \qquad s_{\mu\nu} = \eta_\rho s^\rho{}_{\mu\nu}$$

and using (4.11) one reduces equations (4.8) and (4.9) to

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 8\pi t_{\mu\nu} \tag{4.13}$$

$$Q^\rho{}_{\mu\nu} + \delta^\rho_\mu Q^\sigma_{\nu\sigma} - \delta^\rho_\nu Q^\sigma_{\mu\sigma} = 8\pi s^\rho_{\mu\nu} \tag{4.14}$$

One should note that in the presence of spin and torsion, the tensors $t_{\mu\nu}$ and $R_{\mu\nu}$ need not be symmetric.

4.5. The Bianchi identities lead to

$$Dt_\mu = Q^\nu{}_\mu \wedge t_\nu - \tfrac{1}{2} R^{\nu\rho}{}_\mu \wedge s_{\nu\rho} \tag{4.15}$$

$$Ds_{\mu\nu} = \theta_\nu \wedge t_\mu - \theta_\mu \wedge t_\nu \tag{4.16}$$

where

$$Q^\nu{}_\mu = Q^\nu{}_{\mu\rho}\theta^\rho, \qquad R^{\nu\rho}{}_\mu = R^{\nu\rho}{}_{\mu\sigma}\theta^\sigma$$

The *generalized conservation laws* (4.15) and (4.16) reduce to expected relations in the appropriate limits: (i) If $s_{\mu\nu} = 0$ then $Q^\mu{}_{\nu\rho} = 0$, equation (4.8) or (4.13) reduces to Einstein's, and

$$Dt_\mu = \eta \nabla_\nu t^\nu{}_\mu = 0$$

is the usual differential conservation law of energy–momentum in general relativity theory: (ii) In the limit of special relativity, $R^\mu{}_{\nu\rho\sigma} = 0$, $Q^\mu{}_{\nu\rho} = 0$, one can use rectilinear coordinates (x^μ), and (4.15), (4.16) imply the usual laws,

$$\frac{\partial t^\nu{}_\mu}{\partial x^\nu} = 0 \quad \text{and} \quad \frac{\partial}{\partial x^\rho}(x_\mu t^\rho{}_\nu - x_\nu t^\rho_\mu + s^\rho{}_{\mu\nu}) = 0$$

(iii) If the vector field (v^μ) generates a symmetry of (M, g, ω) then the 3-form

$$v^\mu t_\mu + \tfrac{1}{2}\tilde{\nabla}^\mu v^\nu s_{\mu\nu}$$

is closed.[20] (Here $\tilde{\nabla}$ denotes the covariant derivative with respect to the transposed connection, $\tilde{\omega}^\mu{}_\nu = \omega^\mu{}_\nu + Q^\mu{}_\nu$.) Élie Cartan formulated what is now called the Einstein–Cartan theory as early as in 1923[21] (cf. also the review by Hehl *et al.*,[22] where an extensive bibliography is listed). He demanded, without justification, that the sources satisfy $Dt_\mu = 0$. Together with equation (4.5) this leads to an algebraic relation between curvature and torsion, hard to satisfy otherwise than by assuming $R^\mu{}_{\nu\rho\sigma} = 0$ or $Q^\mu{}_{\nu\rho} = 0$.

4.6. Let (u^μ) be a velocity field, i.e., a smooth vector field on M, normalized by $g_{\mu\nu}u^\mu u^\nu = 1$. Consider the 3-form $u = u^\mu \eta_\mu$ and define, for any tensor field (φ^a) on M, its particle derivative $(\dot{\varphi}^a)$ relative to u,

$$\dot{\varphi}^a \eta = D(\varphi^a u)$$

Following Weyssenhoff and Raabe, a *spinning dust* may be defined as a continuous medium characterized by its velocity (u^μ), the density of energy–momentum (P_μ), and the density of spin $(S_{\mu\nu})$. The 3-forms of energy–momentum and of spin are

$$t_\mu = P_\mu u \quad \text{and} \quad s_{\mu\nu} = S_{\mu\nu} u \tag{4.17}$$

respectively. From equation (4.16) there follows

$$P^\mu = \rho u^\mu - u_\nu \dot{S}^{\nu\mu} \tag{4.18}$$

where

$$\rho = g_{\mu\nu} P^\mu u^\nu$$

Equation (4.16) is equivalent to the system consisting of equation (4.18) and the equation of motion of spin,

$$\dot{S}^{\mu\nu} = u^\mu u_\rho \dot{S}^{\rho\nu} - u^\nu u_\rho \dot{S}^{\rho\mu}$$

whereas equation (4.15) gives rise to the equation of translatory motion

$$\dot{P}_\mu = (Q^\rho{}_{\mu\nu} P_\rho - \tfrac{1}{2} R^{\rho\sigma}{}_{\mu\nu} S_{\rho\sigma}) u^\nu \tag{4.19}$$

which is a generalization, to the Einstein–Cartan theory, of the Mathisson–Papapetrou[46] equation for point particles with an intrinsic angular momentum. From (4.19) it is easy to prove that a spinless test particle moves along a geodesic of the Riemannian connection associated with g, even if M has torsion.

4.7. A simple *cosmological model* may be constructed for a universe filled with a spinning dust.[23] The form (4.17) of energy–momentum and spin density is compatible with the Robertson–Walker line element

$$dt^2 - R(t)^2(dx^2 + dy^2 + dz^2)$$

The field equations reduce to a modified Friedmann equation[24]

$$\left(\frac{dR}{dt}\right)^2 - \frac{2M}{R} + \frac{3S^2}{R^4} = 0 \tag{4.20}$$

where M and S is the total mass and spin within a sphere of radius $R(t)$, respectively. The last term on the left side of equation (4.20) plays the role of a repulsive potential which is effective at small values of R and prevents the solution from ever approaching zero. The significance of this smooth solution is probably restricted because even a small amount of anisotropy and shear will provide a contribution to equation (4.20) reversing the sign of the $1/R^4$ term.[25] In any case, it is clear that the gravitational effects of spin become comparable to those of mass only when the density of spin squared is comparable to the density of energy (in geometrized units these quantities are of the same dimension).

4.8. Let $\mathring{\omega}^\mu{}_\nu$ denote the Levi-Civita (Riemannian) connection form associated with the metric g. Similarly, $\mathring{R}_{\mu\nu}$ and $\mathring{R}$ will be the Ricci tensor and the Ricci scalar of g. The 1-form of the metric connection with torsion $Q^\mu{}_{\nu\rho}$ is given by

$$\omega_{\mu\nu} = \mathring{\omega}_{\mu\nu} + \tfrac{1}{2}(Q_{\mu\rho\nu} + Q_{\rho\mu\nu} + Q_{\nu\mu\rho})\theta^\rho \tag{4.21}$$

If (4.21) is substituted in (4.8) or (4.13) and (4.14) used to solve for torsion in terms of spin, the *Einstein–Cartan equation* (4.13) *assumes the Einstein form*[22]

$$\mathring{R}_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}\mathring{R} = 8\pi(T_{\mu\nu} + \tau_{\mu\nu}) \tag{4.22}$$

where

$$T_{\mu\nu} = t_{\mu\nu} + \tfrac{1}{2}\mathring{\nabla}^\rho(s_{\rho\mu\nu} + s_{\nu\mu\rho} + s_{\mu\nu\rho})$$

is the symmetric energy–momentum tensor obtained from the canonical one, $t_{\mu\nu}$, by the Belinfante–Rosenfeld symmetrization process. The term $\tau_{\mu\nu}$, which is quadratic in $s_{\rho\mu\nu}$, represents the only essential difference between the Einstein–Cartan and the Einstein theories. The difference is not entirely trivial because the canonical tensor, by its relation to the Hamiltonian, may be expected to satisfy "positive energy conditions," whereas there is no clear reason for $T_{\mu\nu} + \tau_{\mu\nu}$ to satisfy any such conditions. In fact, the existence of smooth cosmological models with spin is based on the possibility of circumventing the singularity theorems of Hawking and Penrose by assuming that $t_{\mu\nu}$, rather than $T_{\mu\nu} + \tau_{\mu\nu}$ should satisfy the positivity condition.

4.9. In the past, there was much research and discussion on whether and in what sense gravitation is a gauge theory.[(26–34)] Recently, this problem has been considered in connection with the program of constructing a "supersymmetric" theory of gravitation (cf. References 35–37 and the references given there). In classical relativity, the following questions have been raised and given diverse answers by different authors:

1. What is the gauge group of gravitation?
2. What are the corresponding gauge potentials; what is the status of the metric tensor?
3. Can the form of the field equations be derived from arguments of gauge invariance?

Utiyama was the first to say that gravitation may be looked upon as a gauge theory; he identified its potentials with the coefficients of the Riemannian connection of space–time. Using gauge arguments, Sciama argued in favor of an asymmetric connection as the basis of gravitation and showed that spin may be the source of torsion. Independently, on the ground of heuristic considerations invoking a gauge group with translations (in addition to Lorentz transformations), Kibble derived the full set of field equations of gravitation with spin and torsion; the Sciama–Kibble theory was later recognized as being essentially equivalent to Cartan's theory of 1923; I proposed calling it the Einstein–Cartan theory. Chen Ning Yang pointed out that Einstein's theory is different from other gauge theories in being based on a Lagrangian that is linear, rather than quadratic, in curvature. He proposed considering a theory of gravitation based on Riemannian geometry and a Lagrangian of the form (4.3). The source-free equations of this theory, $\nabla_\mu R_{\nu\rho} = \nabla_\nu R_{\mu\rho}$, appear to be too weak; e.g., they admit as a solution the de Sitter universe with an arbitrary radius of curvature. There is a modification of Yang's theory based on a metric connection with torsion and two sets of field equations, as in the Einstein–Cartan theory. According to Fairchild,[(38)] however, such a Yang–Cartan theory does not have a correct Newtonian limit. Very recently, Hehl, Ne'eman, Nitsch, and von der Heyde[(39)] have formulated a theory of gravitation based on a Lagrangian quadratic in both curvature and torsion. The theory is claimed to have a weak-field limit with a Newtonian and a "confinement" potential, and also an Einstein limit yielding to Schwarzschild solution.

It is clear, from the diversity of results and views, that there is no unique "gauge theory of gravitation." As explained in Sections 1.5 and 4.2, this is due to the fact that gravitation is a "rich" theory from the geometrical point of view: it contains several invariants which may be used to build the kinetic part of the gravitational Lagrangian. The correspondence principle of relativistic gravity to the Newtonian theory suggests—but probably does not require—a Lagrangian linear in curvature, whereas the analogy with electrodynamics leads to the idea of a quadratic Lagrangian. Any theory of gravitation based on a Lagrangian of the latter type requires a careful analysis of its viability.

For me, a gauge theory is any physical theory of a dynamical variable which, at the classical level, may be identified with a connection on a principal bundle. The structure group G of the bundle P is the group of gauge transformations of the first kind; the group $\mathcal{G}$ of gauge transformations of the second kind may be identified with a subgroup of the group Aut P of all automorphisms of P.[40–42] In this sense, gravitation is a gauge theory: the basic gauge field is a linear connection ω (or a connection closely related to a linear connection). In addition to ω, there is a metric tensor g which plays the role of a Higgs field. The most important difference between gravitation and other gauge theories is due to the soldering of the bundle of frames LM to the base manifold M. The bundle LM is constructed in a natural and unique way from M, whereas a noncontractible M may be the base of inequivalent bundles with the same structure group. For example, LS_2 reduced to $SO(2)$ is isomorphic to $SO(3)$, but there is a denumerable set of inequivalent $SO(2)$ bundles over S_2, corresponding to the different elements of $\pi_1(SO(2)) = \mathbf{Z}$. The soldering form θ leads to torsion which has no analog in nongravitational theories. Moreover, it affects the group $\mathcal{G}$, which now consists of the automorphisms of LM preserving θ. This group contains no vertical automorphism other than the identity; it is isomorphic to the group Diff M of all diffeomorphisms of M. In a gauge theory of the Yang–Mills type over Minkowski space–time, the group $\mathcal{G}$ is isomorphic to the semidirect product of the Poincaré group by the group $\mathcal{G}_0$ of vertical automorphisms of P. In other words, in the theory of gravitation, the group $\mathcal{G}_0$ of "pure gauge" transformations reduces to the identity; all elements of $\mathcal{G}$ correspond to diffeomorphisms of M.

What is the structure group G of the gravitational principal bundle P? Since space–time M is four dimensional, if $P = LM$ then $G = GL(4, \mathbf{R})$. But one can equally well take for P the bundle AM of affine frames; in this case G is the affine group. There is a simple correspondence between affine and linear connections which makes it really immaterial whether one works with LM or AM.[7,43] If one assumes—as usually one does—that ω and g are compatible, then the structure group of LM or AM can be restricted to the Lorentz or the Poincaré group, respectively. It is also possible to take, as the underlying bundle for a theory of gravitation, another bundle attached in a natural manner to space–time, such as the bundle of projective frames or the first jet extension of LM.[8] The corresponding structure groups are natural extensions of $GL(4, \mathbf{R})$, $O(1, 3)$ or the Poincaré group.

Acknowledgments

I thank A. Held and J. Ehlers for comments which eliminated an error in Section 1.3 and encouraged me to write Section 4.9.

References

1. T. T. Wu and C. N. Yang, *Phys. Rev. D* **12**, 3845 (1975).
2. N. Steenrod, *The Topology of Fibre Bundles*, Princeton University Press, Princeton, New Jersey (1951).
3. A. Trautman, *Rep. Math. Phys.* (*Toruń*) **1**, 29 (1970).
4. A. Trautman, *Int. J. Theor. Phys.* **16**, 561 (1977).
5. A. Lichnerowicz, *Théorie globale des connexions et des groupes d'holonomie*, Cremonese, Roma (1955).
6. S. Sternberg, *Lectures on Differential Geometry*, Prentice-Hall, Englewood Cliffs, New Jersey (1964).
7. S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Interscience, New York (Vol. I, 1963; Vol. II, 1969).
8. A. Trautman, *Rep. Math. Phys.* (*Toruń*) **10**, 297 (1976).
9. S. Coleman, Classical lumps and their quantum descendants, Lecture at the International School of Subnuclear Physics "Ettore Majorana," unpublished.
10. J. Madore, *Commun. Math. Phys.* **56**, 115 (1977).
11. G. 't. Hooft, *Nucl. Phys.* **B79**, 276 (1974).
12. A. Polyakov, *Zh. Eksper. Teor. Fiz. Pis'ma Red.* **20**, 430 (1974).
13. J. Arafune, P. G. O. Freund, and C. J. Goebel, *J. Math. Phys.* **16**, 433 (1975).
14. S. S. Chern, *Complex Manifolds without Potential Theory*, Van Nostrand, New York (1967).
15. J. Nowakowski and A. Trautman, *J. Math. Phys.* **19**, 1100 (1978).
16. Y. Nambu, Five decades of weak interactions, lecture at International Symposium in Honor of the 60th Birthday of R. E. Marshak, January 1977, City College, New York.
17. A. Trautman, *Ann. N.Y. Acad. Sci.* **262**, 241 (1975).
18. A. Trautman, *Bull. Acad. Polon. Sci., Sér. Sci. Math. Astron. Phys.* **20**, 185, 503, 895 (1972).
19. A. Trautman, in *Ondes et radiations gravitationnelles*, Colloque du CNRS No. 220, CNRS, Paris (1973).
20. A. Trautman, *Bull. Acad. Polon. Sci., Sér. Sci. Math. Astron. Phys.* **21**, 345 (1973).
21. É. Cartan, *Ann. Éc. Norm. Sup.* **40**, 325 (1923).
22. F. W. Hehl, P. Von der Heyde, G. D. Kerlick, and J. M. Nester, *Rev. Mod. Phys.* **48**, 393 (1976).
23. W. Kopczyński, *Phys. Lett.* **A43**, 63 (1973).
24. A. Trautman, *Nature* (*Phys. Sci.*) **242**, 7 (1973).
25. J. Stewart and P. Hajicek, *Nature* (*Phys. Sci.*) **244**, 96 (1973).
26. R. Utiyama, *Phys. Rev.* **101**, 1597 (1956).
27. D. W. Sciama, in *Recent Developments in General Relativity*, Pergamon Press and PWN, Oxford and Warsaw (1962), p. 415.
28. T. W. B. Kibble, *J. Math. Phys.* **2**, 212 (1961).
29. A. M. Brodskii, D. D. Ivanenko, and G. A. Sokolik, *Zh. Eksp. Teor. Fiz.* **41**, 1307 (1961) [*Sov. Phys. JETP* **14**, 930 (1962)].
30. E. A. Lord, *Proc. Camb. Philos. Soc.* **69**, 423 (1971).

31. K. Hayashi and A. Bregman, *Ann. Phys.* (*N.Y.*) **75**, 562 (1973).
32. C. N. Yang, *Phys. Rev. Lett.* **33**, 445 (1974).
33. R. Tabensky, *J. Math. Phys.* **17**, 1 (1976).
34. P. von der Heyde, *Z. Naturforsch.* **31a**, 1725 (1976).
35. S. W. MacDowell and F. Mansouri, *Phys. Rev. Lett.* **38**, 739 (1977).
36. S. Deser, Chapter 12 in this volume.
37. Y. Ne'eman, Chapter 10 in this volume.
38. E. E. Fairchild Jr., *Phys. Rev. D* **14**, 384 (1976).
39. F. W. Hehl, Y. Ne'eman, J. Nitsch, and P. von der Heyde, *Phys. Lett.* **78B**, 102 (1978).
40. W. Drechsler and M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories*, Lecture Notes in Physics No. 67, Springer, Berlin (1977).
41. M. F. Atiyah, N. J. Hitchin, and I. M. Singer, *Proc. R. Soc. London* **A362**, 425 (1978).
42. A. Trautman, *Bull. Acad. Polon. Sci., Sér. Sci. Phys. Astron.* **27**(1), 7 (1979).
43. A. Trautman, *Symp. Math.* **12**, 139 (1973).
44. Y. Aharanov and D. Bohm, *Phys. Rev.* **115**, 485 (1959).
45. C. Ehresmann, Colloque de Topologie (Espaces Fibrés) 5–8 Juin, G. Thone, Liège and Masson, Paris (1951), p. 43.
46. M. Mathisson, *Acta Phys. Polon.* **6**, 163 (1937); A. Papapetrou, *Philos. Mag.* **40**, 937 (1949); *Proc. R. Soc. London* **A209**, 249 (1951).
47. W. Thirring, Gauge theories of gravitation, lecture at XVII International Universitätswochen für Kernphysik, Schladming, 1978.
48. W. Thirring and R. Wallner, *Use of Differential forms in General Relativity*, Vienna University preprint, 1978.
49. B. Carter, *Underlying Mathematical Structures of Classical Gravitation Theory*, Cargese Lecture Notes, 1978.