First Vaidya-Raychaudhuri Endowment Award Lecture

Many Views of the Summit

N. Mukunda

Centre for Theoretical Studies and Department of Physics,
Indian Institute of Science, Bangalore

Indian Association for General Relativity and Gravitation

First Vaidya-Raychaudhuri Endowment Award Lecture

Many Views of the Summit

N. Mukunda

Centre for Theoretical Studies and Department of Physics, Indian Institute of Science, Bangalore

Lecture delivered on November 7, 1989, at Darjeeling during XV Conference of Indian Association for General Relativity and Gravitation organised by the Department of Physics, North Bengal University, West Bengal, India

Introduction

I deem it a great honour and a privilege to be invited to give this talk in honour of Professors Prahlad Chunilal Vaidya and Amal Kumar Raychaudhuri. They are two of the senior and most respected relativists, nay theoretical physicists, in our country today; and this is the first in a series of talks instituted in their names. In a way, through their lifelong devotion to their work, and especially in their choice to fashion their careers and carry out their researches based totally in India, they remind us of Satyendranath Bose and Meghnad Saha, theorists and pioneers of an earlier generation. It is with respectful admiration that I present this talk to them.

I would like to address myself to students and young research workers getting interested in problems in relativity and gravitation, and to present with a light touch some points of view illuminating different aspects of the general theory of relativity. This theory is legitimately regarded as the summit of classical, i.e. pre-quantum, physics; and is often described as a supreme combination of physical insight, mathematical beauty and harmony, "one of the greatest examples of the power of speculative thought". Indeed in the words of Landau and Lifshitz, "It is probably the most beautiful of all existing theories". It is instructive and important to realise, however, that the triumphant completion of the general theory of relativity by Albert Einstein in November 1915 was preceded by a long period of gestation, several years of hard labour, with quite a few false steps and misconceptions on the way. Especially in view of some remarks to be made later on, let us rapidly remind ourselves of some parts of this heroic one-man struggle.

The Meaning of General Covariance

As we all know, the essential structure of special relativity had been worked out by Einstein, sitting in his chair in the patent office in Berne, in 1905. It was in November 1907, sitting in that same chair preparing a review of special relativity, that there occurred the first of many deep insights into the problem of gravitation[1]. Later characterised by him as "the happiest thought of my life", it was the first understanding of the Principle of Equivalence, that a gravitational field could be eliminated or transformed away by passing to a suitably accelerated noninertial frame of reference. In attempting to reconcile Newtonian gravity and special relativity, Einstein saw *very soon* that in fact special relativity had to be *extended* to include gravity. With great courage, after this realisation he gave up Lorentz invariance as a global requirement, and in a manner of speaking began the search for a "larger symmetry group" to encompass gravitation. During the period April 1911 to August 1912 spent at Prague, several important advances were made: first analysis of the gravitational bending of light; the gravitational red shift; the dependence of the speed of light on the gravitational field; and probably more important than all these, the understanding on the one hand that Newtonian gravitation and special relativity were both incomplete, and on the other hand that the equivalence principle is only a local statement. Added to all these, Einstein began to take the first steps on the long road to finding a dynamical description of gravitation itself.

By the time Einstein moved to Zurich in August 1912, the necessity of giving up Euclidean geometry, of enlarging Lorentz invariance to covariance under general coordinate transformations, and also of replacing the single scalar Newtonian gravitational potential by a ten-component metric tensor field, had all become clear. Then he suddenly realized that the key to his problems lay in the Gauss' theory of surfaces developed in the early 19th century. His friend Marcel Grossmann introduced him to the later work of Riemann, and then on to the Italian masters

Gregorio Ricci-Curbastro and Tullio Levi-Civita. The problem posed by Einstein to Grossmann was: use the theory of invariants and covariants under general coordinate transformations to construct a suitable tensor $\Gamma_{\mu\nu}$ out of the metric tensor $g_{\mu\nu}$ and its spacetime derivatives; thus arrive at a generalisation of the Poisson equation for the Newtonian gravitational potential in the symbolic form:

$$\nabla^2\phi = 4\pi G\rho \rightarrow \Gamma_{\mu\nu}\,(g,\, \frac{\partial g}{\partial x}) = K\theta_{\mu\nu},$$

$\theta_{\mu\nu}$ = matter-radiation energy momentum tensor,

$$K = 8\pi G/c^4. \tag{1}$$

However at this stage Einstein had not yet fully grasped the extremely subtle implications of the requirement of general covariance! On incorrect physical reasoning he had convinced himself that the field equations he was looking for must have the property that, given appropriate boundary conditions, the ten metric functions $g_{\mu\nu}(x)$ on spacetime ought to be *completely determined* by a given source function $\theta_{\mu\nu}(x)$. In today's terminology, Einstein and Grossmann had not appreciated the gauge aspect of the problem, the fact that the freedom to perform arbitrary general coordinate transformations should legitimately leave the field equations *underdetermined* to that extent, and that this was to be expected. They were at this stage deceived by the covariant constancy of the metric,

$$g_{\mu\nu\,;\lambda} = 0, \tag{2}$$

and were unaware of the Bianchi identities. Due to these reasons, in their 1913 work they deliberately reduced the earlier postulated general covariance under *all* coordinate transformations to invariance under *linear* transformations alone, definitely a backward step.

One can understand in retrospect what a struggle it must have meant to think through and analyse such delicate problems, something never before encountered in any physical theory. It recalls to mind Galileo's struggles three centuries earlier to arrive at the most useful definition of acceleration. The resolution came during the period July to November 1915. While the demand of general covariance under curvilinear coordinate transformations was reinstated, for a while Einstein limited himself to unimodular transformations for which the Jacobian was unity - we see it as the desire to avoid distinguishing scalars and tensors from the corresponding densities. At a later stage even the demand

$$det\left(g_{\mu\nu}(x)\right) = 1 \tag{3}$$

on the metric was imposed. The final resolution of all these problems, and the elucidation of the field equations in complete form, came by November 25, 1915: even at this stage, though, the Bianchi identities were not in his grasp! While the field equations had essentially been cast into the form

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -K\theta_{\mu\nu}, \tag{4}$$

the covariant conservation of the left hand side was regarded as a constraint on it imposed by the conservation of the energy momentum tensor:

$$\theta^{\mu\nu}_{;\nu} = 0 \Longrightarrow \left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)_{;\nu} = 0. \tag{5}$$

Today of course we do just the opposite! The right hand side is an identity, so for consistency the sources must obey the left hand side.

Both the joy of discovery of the final result, and the clearing away of the misconceptions of so many years of effort, are best captured in Einstein's own words of 1915: "Scarcely anyone who fully understands this theory can escape from its magic", and much later in 1933: "The years of searching in the dark for a truth that one feels but cannot express, the intense desire and the alternations of confidence and misgiving until one breaks through to clarity and understanding are known only to him who has himself experienced them".

The Action for Gravitation

Just like Newton's enunciation of the principles of dynamics, and later Maxwell's equations for the electromagnetic field, here too the new law was discovered directly at the level of differential equations of motion. The idea of a variational principle and an action came in each case as a later formal development. In the case of general relativity, however, it was a very close thing indeed[2]. Einstein and the mathematician David Hilbert had been in frequent correspondence, especially during November 1915. And it happened that on November 20th, five days before Einstein presented his final field equations, Hilbert submitted a paper containing essentially the same equations but derived from a variational principle. It is sobering to know, however, that he too was then unaware of the Bianchi relations! Thus Hilbert was the first to give the correct expression for the action in general relativity. One important difference was that while Einstein had left unspecified the expression for the energy-momentum tensor $\theta_{\mu\nu}$, except for its symmetry and covariant conservation, Hilbert had assumed a definite action for the sources too and so had specific expressions for $\theta_{\mu\nu}$. Initially there was a slight misunderstanding, essentially on grounds of priority, but it was evidently later cleared up between them. But from a human point of view it is interesting, indeed reassuring, to see evidence of attachment to one's proudest discoveries. Around that period, Einstein said of Hilbert's work: "Hilbert's ansatz for matter seems childish to me"; and on another occasion a little later, "I don't like Hilbert's presentation . . . unnecessarily special . . . unnecessarily complicated . . . not honest in structure". One may well compare such expressions of feeling with what Heisenberg and Schrödinger said of one another's work a decade later[3]. In a footnote to his paper establishing the equivalence of matrix and wave mechanics, Schrödinger says: "My theory was inspired by L. de Broglie and by brief but infinitely far-seeing remarks of A. Einstein. I was absolutely unaware of any genetic relationship with Heisenberg. I naturally knew about his theory, but because of the very difficult appearing methods of transcendental algebra and the lack of Anschaulichkeit, I felt deterred by it, if not to say repelled". But Heisenberg was not to be out done, as is evident from his letter to Pauli: "The more I think of the physical part of the Schrödinger theory, the more abominable I find it. What Schrödinger writes about Anschaulichkeit makes scarcely any sense, in other words I think it is bullshit. The greatest result of his theory is the calculation of the matrix elements". What human sentiments and expressions! May be in German it all sounds quite harmless!

Constrained Dynamics, Reduction of Manifest Covariance

Let me now return to one of my main themes. The importance of the action for a dynamical system, as standing logically prior to classical equations of motion, was emphasized by the development of quantum theory. An action based on a Lagrange function of course also allows passage to the canonical Hamiltonian form, as a step in the quantisation process. One of the first attempts to give a canonical treatment for general relativity is due to Leon Rosenfeld, in 1930[4] .

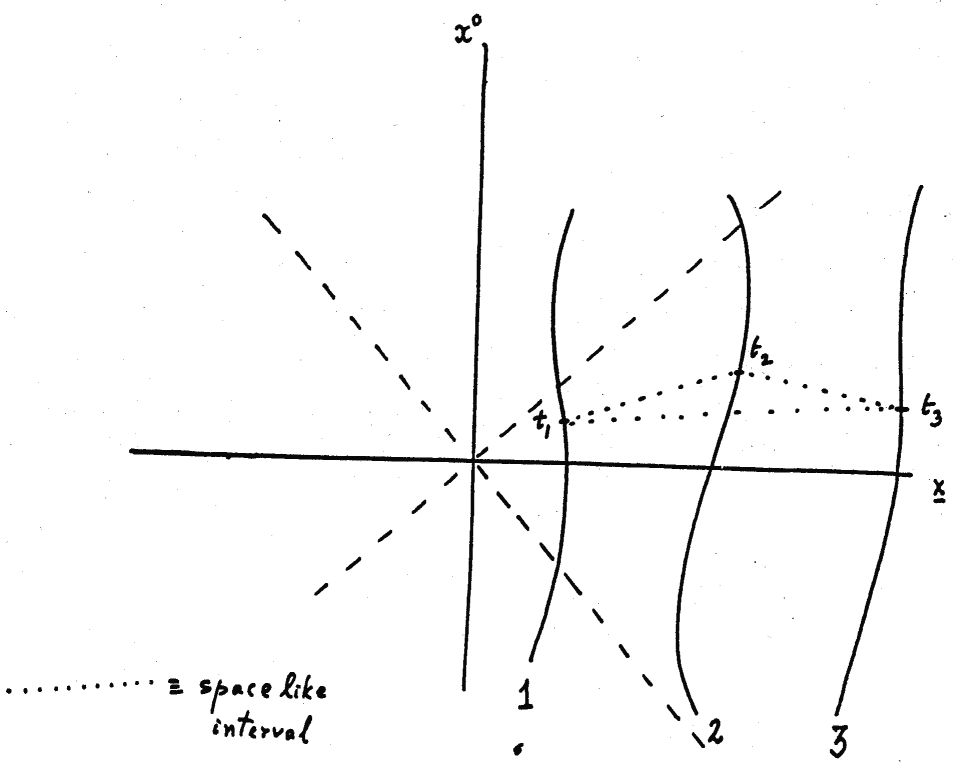
But he ran into difficulties in applying the standard rules familiar from mechanics, again for the same technical mathematical reasons that had caused Einstein so much trouble earlier. Namely, because of the general covariance under arbitrary coordinate changes, the field equations do not and cannot determine an unambiguous time evolution for all ten metric field components. During the thirties Dirac did some work of a general nature concerning such dynamical systems for which the Lagrangian fails to permit a simple passage to the Hamiltonian formalism[5]. Specifically he studied the properties of systems for which the Lagrangian is homogeneous of degree one in velocities:

$$\dot{q}^j \frac{\partial L(q, \dot{q})}{\partial \dot{q}^j} = L(q, \dot{q}). \tag{6}$$

In such cases he introduced the phrase "Hamiltonian equation" in place of "Hamiltonian function": as one sees, the naive Hamiltonian simply vanishes. Starting sometime in the early forties, Dirac then undertook a systematic and important extension of the age-old formalism of classical dynamics, creating new concepts and methods capable of handling such and other singular Lagrangian systems. The subject goes by the name "Constrained Hamiltonian Dynamics", and his final comprehensive presentation of it was at a Canadian mathematical seminar in 1949[6]. It is indeed a very beautiful and powerful contribution to the general formalism of analytical dynamics, which reveals its true value in relativistic problems and more recently in the analysis of gauge theories. The strongest motivation for this work was of course the desire to give a satisfactory and complete canonical treatment of general relativity. Around the late forties and early fifties Peter Bergmann and his collaborators also did some very original work analysing canonical aspects of generally covariant field theories, and the concepts of primary and secondary constraints are due to them[7]. But the setting up of a comprehensive formalism for constrained systems, with the additional notions of first and second class constraints, their distinct transformation theoretic roles, and the important concept of the Dirac bracket, are all due to Dirac. The application of these methods to the Einstein-Hilbert action for general relativity was given by him in 1958[8], while Arnowitt, Deser and Misner also gave a comparable treatment soon after[9].

The importance of the canonical formalism lies mainly in the fact that the true dynamical degrees of freedom are unambiguously identified and separated from nondynamical variables. Dirac was able to show in a clean way that the four components $g_{\mu o}$ of the metric drop out as being essentially nondynamical, corresponding to the four-fold freedom of general coordinate transformations. He found a way to alter the Einstein-Hilbert action density, amounting to a contact transformation, to make this quite explicit. However the use of canonical methods generally reduces the extent to which manifest covariance can be maintained. Already one restricts the possible choices of spacetime coordinate systems so that each three-dimensional surface $x^o = constant$ is spacelike; and then one limits general coordinate transformations to those that maintain this property. As a result of referring the dynamics to these special surfaces, the manifest general covariance of the action or Lagrangian based treatment is considerably reduced. At the end of his efforts in 1958 Dirac was led to say[8]: "One starts with ten degrees of freedom for each point in space, corresponding to the ten $g_{\mu\nu}$, but one finds with the method here followed that some drop out, leaving only six, corresponding to the six g_{rs}. This is a substantial simplification, but it can be obtained only at the expense of giving up four-dimensional symmetry. I am inclined to believe from this that four-dimensional symmetry is not a fundamental property of the physical world.... The present paper shows that Hamiltonian methods, if expressed in their simplest form, force one to abandon the four-dimensional symmetry". Somewhat later in 1962 he expressed it thus :[10] "The gravitational field is a tensor field with ten components. One finds that six of the components are adequate for describing everything of physical importance and the other four can be dropped out of the equations. One cannot, however, pick out the six important components from the complete set of ten in anyway that does not destroy the four-dimensional symmetry. Thus if one insists on preserving four-dimensional symmetry in the equations, one cannot adapt the theory of gravitation to a discussion of measurements in the way quantum theory requires without being forced to a more complicated description than is needed by the physical situation. This result has led me to doubt how fundamental the four-dimensional requirement in physics is". Here one may mention that Regge and collaborators have recently tried to build up a formalism which combines the virtues of differential geometric methods - which capture the essence of general covariance - with the spirit of Hamiltonian methods, to ameliorate the situation described by Dirac[11].

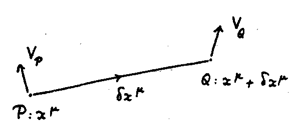
The use of general curved space like three-dimensional surfaces in spacetime for specifying initial data in a canonical framework, and the reduction in the extent of manifest covariance when one uses canonical methods, are both already familiar within special relativity. In passing from the Galilean to the special relativistic view of spacetime, the most important change is in the meaning of simultaneity. The set of events which can be regarded as being simultaneous with a given event gets enlarged from merely a three- dimensional (flat) *section* to a non trivial (open) four-dimensional *region* in spacetime. This leads to far greater flexibility in viewing a relativistic dynamical problem as the "evolution of initial data". Already in 1932 Dirac, Fock and Podolsky built up the so-called multi-time formalism to describe a relativistic many-electron theory[12]. Thus, instead of choosing a single common time (in some inertial frame) as the evolution parameter, one picks instead a separate time parameter $t_1, t_2, \ldots$ for each of the electrons; one must only ensure that (classically!) the corresponding spacetime points on the various world-lines are pairwise mutually space like.



This formalism led later to the Tomonaga-Schwinger formalism in special relativistic quantum field theory: for a general curved space like hypersurface σ in Minkowski spacetime, one has a state vector (functional) $\Psi[\sigma]$ which is subject to a functional Schrödinger equation under independent local variations of σ[13]. And one can view the canonical formalism of general relativity as the end product of this line of development. Some ways of presenting canonical special relativistic theories, each of which takes advantage of the enlarged region of simultaneity in a particular way and so singles out a particular subgroup of the Poincaré group as a manifest symmetry, have also been given by Dirac[14]. These are the so-called instant, point and front forms. The first is the familiar form in which just the six-parameter Euclidean group on space appears as a manifest symmetry. In the point and front forms, however, the manifest symmetry is with respect to the homogeneous Lorentz group, and an "unusual" seven-parameter subgroup of the Poincaré group respectively. These correspond in turn to choosing initial data on a positive time like hyperboloid with respect to a given spacetime point, which is a space like surface; and on a light like hyperplane which is (almost) space like.

The Gauge Idea, Fibre Bundle Methods

Let us now turn to another line of development, leading to a different view of the summit. This has to do with efforts to modify and enlarge the geometric foundations of general relativity, to encompass other physical fields of force or fundamental interactions. The first major attempt in this direction, due to Hermann Weyl, was in 1918, and this is the origin of the gauge idea[15]. The aim was to unify gravitation and electromagnetism in a single comprehensive geometrical framework. Let us look briefly at the main ideas. In Einstein's theory, based on a (pseudo) Riemannian metric and accompanying geometry, there is a definition of parallel transport of vectors and tensors given by the Christoffel connection. This is symmetric and is fully determined by the metric. Lengths and angles among vectors are preserved under this parallel transport, so that the metric itself is covariantly constant. However after parallel transport over a closed circuit a vector may end up with an altered direction, which then signals the existence of nonzero curvature. Weyl extended this to allow for lengths also to change under parallel transport. He assumed that parallel transport was given by a *symmetric* affine connection; this amounts to saying that one can choose geodesic coordinates around each space-time point, or equally well that the torsion vanishes. But under parallel transport lengths of (contravariant) vectors as determined by a metric $g_{\mu\nu}(x)$ could change, while angles were still preserved. Parallel transport over a closed circuit could then result in a vector changing both direction and length. If a vector V_P^μ located at a space time point P with coordinates x^μ is transported to a nearby point Q with coordinates $x^\mu + \delta x^\mu$, we get at Q a vector V_Q^μ:



$$V_Q^\mu = V_P^\mu - \Gamma^\mu_{\rho\sigma}\, \delta x^\rho V_P^\sigma,$$

$$\Gamma^\mu_{\rho\sigma} = \Gamma^\mu_{\sigma\rho}. \tag{7}$$

As angles are to be preserved, the length change suffered by V_P must be by a fractional amount which is itself V_P - independent. Therefore there must be a linear form in δx^ρ such that

$$g_{\mu\nu}(Q)V_Q^\mu V_Q^\nu = (1 - A_\rho(x)\delta x^\rho) g_{\mu\nu}(P) V_P^\mu V_P^\nu. \tag{8}$$

So the connection coefficients are not just the metric compatible Christoffel coefficients $\Gamma^{(0)}$, there is an additional part coming from the vector field A_μ :

$$\Gamma^\lambda_{\mu\nu} = \Gamma^{(0)\lambda}_{\mu\nu} + \frac{1}{2}\left(\delta^\lambda_\mu A_\nu + \delta^\lambda_\nu A_\mu - g_{\mu\nu} A^\lambda\right),$$

$$\Gamma^{(0)\lambda}_{\mu\nu} = \text{Christoffel connection coefficients.} \tag{9}$$

(Weyl's geometry thus accommodates both a metric tensor field $g_{\mu\nu}(x)$ and a fundamental vector field $A_\mu(x)$). Correspondingly, with respect to the covariant differentiation defined by Γ, the metric is *not* constant:

$$D_\lambda\, g_{\mu\nu} \equiv (\partial_\lambda + \Gamma^{\cdot}_{\lambda\cdot} + \Gamma^{\cdot}_{\lambda\cdot}) g_{\cdot\cdot} \neq 0. \tag{10}$$

Since lengths are not preserved under parallel transport, in Weyl's geometry we have the freedom to change the unit of length independently at each spacetime point. Thus, compared to the Riemannian case, we now have both the freedom to perform general changes of coordinates and to perform "changes of gauge":

$$\begin{aligned} g_{\mu\nu}(x) &\rightarrow g'_{\mu\nu}(x) = e^{\sigma(x)} g_{\mu\nu}(x), \\ \Gamma^\lambda_{\mu\nu}(x) &\rightarrow \Gamma'^{\lambda}_{\mu\nu}(x) = \Gamma^\lambda_{\mu\nu}(x), \\ \mathrm{V}^\mu(\mathrm{x}) &\rightarrow \mathrm{V}'^\mu(\mathrm{x}) = \mathrm{V}^\mu(\mathrm{x}) \end{aligned} \tag{11}$$

The last two are natural assumptions, and consistency then gives,

$$A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) - \partial_\mu \sigma(x). \tag{12}$$

On this basis Weyl tried to identify $A_\mu(x)$ with the vector potential of electromagnetism. The field strength

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x), \quad (13)$$

is of course gauge-invariant; and if it is nonzero it indicates that after parallel transport over a closed circuit the length of a vector definitely changes.

Each significant field $X(x)$ in Weyl's theory carries a weight n, the integer appearing in its gauge transformation law:

$$X(x) \rightarrow X'(x) = e^{n\sigma(x)} X(x). \quad (14)$$

(Exceptions, such as $A_\mu(x)$ which has a linear inhomogeneous law of change, would be obvious). Thus for example we have

$$n = 1 : g_{\mu\nu}(x), \;\; \mathrm{V}_\mu(\mathrm{x}), \;\; \Gamma_{\lambda\mu\nu}(\mathrm{x});$$

$$n = 0 : \mathrm{V}^\mu(\mathrm{x}), \;\; \Gamma^\lambda_{\mu\nu}, \;\; \mathrm{R}^\lambda_{\mu\nu\rho}(\mathrm{x}), \;\; \mathrm{R}_{\mu\nu}(\mathrm{x});$$

$$n = -1 : g^{\mu\nu}(x), \;\; R(x). \quad (15)$$

Here $R^\lambda_{\mu\nu\rho}$ is the curvature tensor formed out of $\Gamma^\lambda_{\mu\nu}$, $R_{\mu\nu}$ the corresponding Ricci tensor, and R the curvature scalar. A combined gauge and coordinate covariant derivative ∇_μ can be set up, symbolically

$$\nabla_\mu = \partial_\mu + \Gamma^{\cdot}_{\mu\cdot} + nA_\mu, \quad (16)$$

and then the metric obeys

$$\nabla_\lambda \, g_{\mu\nu} = 0. \quad (17)$$

There are tell-tale differences between a Weyl-type theory and conventional general relativity: geodesics are no longer lines of extremal path length since the former are determined by $\Gamma^{\lambda}_{\mu\nu}$ and the latter by $g_{\mu\nu}$; and the Lagrange density which has to be a scalar density of gauge weight $n = o$ can no longer be taken to be $\sqrt{-g}R$ whose weight is $n = 1$. Thus the gravitational field equations are necessarily quite different from the Einstein equations. Some candidate expressions for the gravitational action density are

$$\sqrt{-g}(R^{\lambda\mu\nu\rho}R_{\lambda\mu\nu\rho} \ \ or \ \ R^{\mu\nu}R_{\mu\nu} \ or \ R^2). \tag{18}$$

Nevertheless, it can be arranged that the perihelion shift and bending of light are the same as in ordinary general relativity.

It was however soon pointed out by Einstein that this scheme was physically untenable. Linking up gauge changes in the electromagnetic vector potential with real local changes in time and length units would mean that different chemical elements would not have definite sharp and characteristic spectral lines at all. But the idea that such changes in this vector potential should be the accompaniment of some other physically important transformation in some other significant quantity survived. Indeed inspite of its being untenable in the specific form in which he first proposed it, Weyl felt "it was so beautiful that he did not wish to abandon it and so he kept it alive for the sake of its beauty". Almost a decade later, the gauge idea for electromagnetism found its proper expression after the discovery of wave mechanics by Schrödinger in 1926. Namely V. Fock[16] and F. London[17] in 1927, and Weyl himself in 1929[18], saw that by combining the classical canonical rule

$$p_\mu \ \rightarrow \ p_\mu - \frac{e}{c}A_\mu(x) \tag{19}$$

for coupling a charged point particle to the electromagnetic field, with the wave-mechanical prescription for the energy-momentum operator

$$p_\mu \rightarrow \frac{\hbar}{i}\frac{\partial}{\partial x^\mu} \tag{20}$$

designed to act on the complex wave function $\psi(x)$ of a point particle, one obtained

$$\frac{\hbar}{i}\frac{\partial}{\partial x^\mu} \rightarrow \frac{\hbar}{i}\frac{\partial}{\partial x^\mu} - \frac{e}{c}A_\mu(x). \tag{21}$$

for coupling a charged quantum mechanical particle to the electromagnetic field. This is the origin of the concept of "minimal electromagnetic coupling". In contrast to the doubly covariant derivative ∇_μ in Weyl geometry, here $\partial/\partial x^\mu$ and $A_\mu(x)$ combine with a relative pure imaginary, rather than real, coefficient. This then meant that the gauge change (12) in the vector potential goes with a *phase* change in the complex wave function of a charged particle:

$$\psi(x) \rightarrow \psi'(x) = e^{\frac{ie}{\hbar c}\sigma(x)}\psi(x). \tag{22}$$

And the combination (19) of $\partial/\partial x^\mu$ and $A_\mu(x)$ enjoys the property

$$\left(\frac{\partial}{\partial x^\mu} - \frac{ie}{\hbar c}A'_\mu(x)\right)\psi'(x) = e^{\frac{ie}{\hbar c}\sigma(x)}\left(\frac{\partial}{\partial x^\mu} - \frac{ie}{\hbar c}A_\mu(x)\right)\psi(x). \tag{23}$$

Thus instead of viewing the gauge transformations of electromagnetism as an expression of the noncompact group R of real numbers under addition, it is physically to be seen as an expression of the compact group U(1) of complex quantum mechanical phase factors[19]. Indeed even the gauge transformation law (12) for $A_\mu(x)$ can be expressed via such phase factors:

$$A'_\mu(x) = e^{\frac{-ie}{\hbar c}\sigma(x)}\left(A_\mu(x) - \frac{\hbar c}{ie}\frac{\partial}{\partial x^\mu}\right)e^{\frac{ie}{\hbar c}\sigma(x)} \tag{24}$$

While this is still a local transformation, varying from point to point in spacetime, the action itself is not on spacetime or on the metric, but in an internal U(1) space attached to each point of space- time.

The group U(1), besides being compact, is also Abelian. Over the ensuing decades the highly nontrivial generalization to a compact but nonabelian internal gauge symmetry group G was achieved, starting with Oskar Klein in 1938[20] and independently rediscovered by C. N. Yang and R.L. Mills in 1954[21]. The motivations came from nuclear and elementary particle physics. Here one envisages a "copy" of the group G, or in some circumstances the space V of a representation D(g) of G, "attached" to each point of spacetime; and then one has a multicomponent field $\psi(x)$ belonging to this representation space. One also has a nonabelian generalization of the real vector potential of electromagnetism to a vector potential $A_\mu(x)$ which is simultaneously a hermitian matrix in the Lie algebra of G (corresponding, say, to the representation D(g)), and a covariant vector on spacetime. And when one subjects $\psi(x)$ to a spacetime dependent transformation of G, A_μ changes in a linear inhomogeneous fashion such that the combination $\partial_\mu - ieA_\mu$ again behaves nicely:

$$\psi'(x) = D\Big(g(x)\Big)\psi(x),$$

$$A'_\mu(x) = D\Big(g(x)\Big)\Big(A_\mu(x) + \frac{i}{e}\partial_\mu\Big)D\Big(g(x)\Big)^{-1},$$

$$\Big(\partial_\mu - ieA'_\mu(x)\Big)\psi'(x) = D\Big(g(x)\Big)\Big(\partial_\mu - ieA_\mu(x)\Big)\psi(x). \tag{25}$$

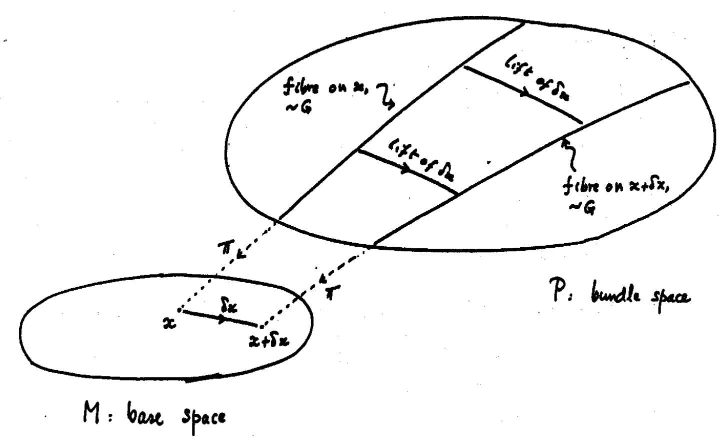
There is also a nonabelian field strength tensor $F_{\mu\nu}(x)$, again a hermitian matrix in the Lie algebra of G, which unlike the electromagnetic case is gauge covariant rather than gauge invariant:

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) - ie[A_\mu(x),\ A_\nu(x)],$$

$$F'_{\mu\nu}(x) = D\Big(g(x)\Big) F_{\mu\nu}(x) D\Big(g(x)\Big)^{-1}. \qquad (26)$$

The nonlinearities in these expressions as compared to electromagnetism, when G = U(1), are evident. Indeed the nonabelian A_μ and $F_{\mu\nu}$ here remind us of the connection Γ and curvature R on an affinely connected manifold or on a Riemannian manifold!

The proper mathematical framework for viewing these structures turns out to be the theory of fibre bundles, more especially principal fibre bundles[22]. Here one has a base manifold M, for instance spacetime, and a copy of a Lie group G attached to each point of M, as a fibre sitting "on top of" that point, in a smooth way. Then the total space P looks locally like the Cartesion product of a portion of M with G, but globally it need not have the character of a product at all. In fact the principal fibre bundle is nontrivial precisely to the extent to which it is globally distinct from M × G. What is globally given, as schematically indicated in the figure, is a projection map $\Pi : P \rightarrow M$.



which takes all the points on each fibre to the base point in M "under the fibre". Also globally given is an action of G on each fibre, which is both free and transitive; recalling that each fibre is essentially G itself, this action is locally realised as, say, right translation using the group composition law of G. In this set-up, the gauge or vector potential A_μ is a local description of a *connection* on P: a rule for parallel transport. A connection on a principal fibre bundle is a quite delicate and beautiful concept - for each small displacement $x \rightarrow x + \delta x$ in the base manifold M, it gives a definite rule to move from any point in the fibre on top of the point x to a definite corresponding point in the fibre on top of $x + \delta x$. Remember here that each fibre looks like and can be taken to be the group G itself. Thus a connection tells us how to "lift" an open or closed curve in the base M to definite curves in the total space P, and these lifted curves go into one another under the global G action on the fibres. However the lifts of a closed curve in M could well be open in P: one returns to the same fibre but to a generally different point than the starting one! This signals the existence of and is described by the curvature, of which $F_{\mu\nu}$ is a local expression.

The theory of principal fibre bundles and connections on them was set up by Cartan and others quite early, in the twenties and thirties. (It was around the same time that Cartan developed the elegant calculus of differential forms). The nice thing in this framework is that the linear inhomogeneously transforming vector potential A_μ defined over (local portions of) the base manifold M is elevated to a covariantly behaving geometrical object - a connection form - on the *total* space P. Both A_μ and $F_{\mu\nu}$ on M are local descriptions via local cross-sections of G-covariant and globally defined geometrical objects more properly viewed as existing in the larger space P. Then the gauge transformation rules for A_μ and $F_{\mu\nu}$ are results of changing this local description by changing the cross-section.

In this language one can say that while Weyl's original theory attempted to view electromagnetism via an R-bundle over spacetime M, each fibre being essentially the entire real line viewed as a group under addition, later developments in quantum mechanics replaced this by a U(1) - bundle over spacetime. At the same time, Weyl's attempt to modify Riemannian geometry got transmuted to a scheme in which the gauge transformations of electromagnetism are tied up with transformations in an *internal* space, referring to a non-spacetime symmetry. How does all this now connect up with general relativity ? Can we with this additional insight and the later development of nonabelian gauge theory go back and reexamine general relativity itself, and view it as a kind of gauge theory ? Here the story goes back again to Cartan and his work in the twenties![23] While a given base manifold M (of dimension n, say) and a given Lie group G can in principle be combined in many inequivalent ways to form many principal fibre bundles, and one could then conceive of connections on each of them, there is a particular case which is specially singled out: a given base space M *on its own determines unambiguously* a particular principal fibre bundle for which the fibre or structure group is $GL\ (n,\ R)$. For $n = 4$ we have $GL\ (4,\ R)$. The point is that *there is no freedom* in the way in which M and $GL\ (n,\ R)$ are glued together to form this particular bundle, the so-called *frame bundle on M;* it is completely fixed by the global topological properties of M itself. The key concept here is Cartan's "repere mobile": a freely chosen anholonomic non-coordinate based frame or basis of vectors at each point of M. In four dimensions we call it a vierbein or tetrad. Even without a metric on M we have the framework of this specific principal fibre bundle in which affine connections and parallel transport of vectors and tensors can be thought of : we need to deal with connections on the frame bundle of M. In a local description this connection is given by affine connection coefficients $\Gamma^{\lambda}_{\mu\nu}(x)$ over M, which are in general not symmetric, so there is room for torsion. Thus while in a general principal fibre bundle with a group G over a base M the key concepts are just those of connection and curvature only, in the particular case of the frame bundle we have connection, curvature and torsion (not to speak of torture) as well. Another way of singling out this particular principal fibre bundle is to say that with the help of the dual basis to the vierbein, which constitutes a covariant geometric object or form called the "soldering form", the individual fibres $GL(4, R)$ are soldered to points of the base M in a more intimate way than the fibres of a general principal fibre bundle. There is extra geometrical structure in a frame bundle. All this was worked out by Cartan in complete generality by about 1922! In case M carries a (pseudo) Riemannian metric as well, the tetrad can naturally be chosen to bring the metric $g_{\mu\nu}$ at each point to diagonal Minkowski form, and then the Lorentz group SO(3,1) appears as a local gauge group as well.

That the fibre bundle concept should be useful to handle a spacetime can be seen in another more elementary sense as well. In fact this becomes evident in the passage from the Aristotelian to the Galilean principle of relativity[22]. In Aristotelian physics governed by the notions of absolute time and absolute space or absolute rest, each of the two statements - "the events A and B are simultaneous in time", "the events A and B occur at the same spatial location at different times" - has an invariant meaning. Thus spacetime is just the Cartesian product of space and time. But with Galilean relativity where two inertial observers could have a uniform relative velocity, the second statement above no longer has an invariant meaning ! Simultaneity in time still has an absolute meaning but not so coincidence in space. Therefore spacetime is not a Cartesian product any more. Since time differences between events have absolute meaning we can say that what is well-defined is a projection from spacetime on to the one- dimensional time axis, so the former is a fibre bundle over the latter as base, with each fibre isomorphic to three-dimensional space. With special relativity, of course, there are further changes and this bundle structure is lost.

One can now try to close the circle and ask: are geometrical theories of gravity also gauge theories, similar to Yang-Mills theories based on an internal symmetry? The answer is a qualified yes, to some extent dependent on definitions of terms. An attempt to answer this question was first made by Utiyama in 1956[24] and completed more satisfactorily by Kibble in 1961[25] and Sciama in 1962[26]. The situation is that if one starts with a special relativistic Lagrangian field theory invariant under global Poincare' transformations, and in the spirit of the Yang-Mills argument one makes the ten parameters of the Poincaré group arbitrary independent spacetime functions, one can then in a natural and intelligent way modify the Lagrangian so as to be invariant under this gauged Poincaré group. The resulting theory is then a generally covariant field theory of the same general type that Cartan and invented in the 1920's! This is called the Einstein-Cartan approach to gravity, on account of an interesting correspondence between them in the period 1929-1932 when Einstein was working with the idea of absolute or distant parallelism and Cartan told him, as usual, that he had done it years ago[27]. Adopting the Yang-Mills method from an internal compact symmetry group G to the Poincaré group, one naturally finds that several geometric objects have to come in: a vierbein or tetrad h on spacetime; a "spin-connection" A which is just like a Yang-Mills gauge potential but with the Lorentz group SO(3,1) (more properly SL(2,C)) as a local internal gauge group; and an affine connection Γ on spacetime:

$$h^{\mu}_{a} = \textit{vierbein or tetrad};$$

$$h^a_\mu \; = \; inverse \; to \; tetrad \; = \; components \; of \; solderingform;$$

$$A^{ab}_\mu \; = \; -A^{ba}_\mu \; = \; spin \; connection \; =$$

$$Yang - Mills \; gauge \; potential \; for \; local \; SL \; (2,C) \; transformations;$$

$$\Gamma^\lambda_{\mu\nu} \; = \; affine \; connection \; coefficients \; on \; spacetime. \qquad (27)$$

One is dealing here with a principal fibre bundle on spacetime which is a combination of the frame bundle and an SL(2,C) bundle! There is thus room here for both curvature and torsion, unlike a pure internal gauge symmetry; this agrees with what we said earlier, namely the frame bundle on a base manifold has richer structure and more intrinsic geometric objects than a general principal fibre bundle. Giving due allowance for these facts, one can view the Einstein-Cartan theory as the result of gauging the Poincaracutee group of special relativity. As the vierbein has one leg in each group - GL(4,R) and SL(2,C) - contained in the fibres, the natural condition to impose is

$$\nabla_\mu h^\nu_a \; \equiv \; (\partial_\mu \; + \; \Gamma^{\cdot}_{\mu\cdot} \; + \; A^{\cdot\cdot}_\mu) h^{\cdot}_{\cdot} \; = \; 0. \qquad (28)$$

This is like the square root of the covariant constancy of the metric in ordinary general relativity! The vierbein and the spin connection can be regarded as the fundamental gravitational variables. In terms of them, both the metric and the affine connection, and then the torsion and the curvature, are determined:

$$g_{\mu\nu} \; = \; h_{\mu a} \, h^a_\nu;$$

$$\Gamma^{\lambda}_{\mu\nu} = h^{\lambda}_{a} D_{\mu} (A) h^{a}_{\nu},$$

$$D_{\mu}(A) = SL(2,C) - \textit{covariant derivative};$$

$$T^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\mu\nu} - \Gamma^{\lambda}_{\nu\mu} = \textit{torsion};$$

$$R^{\lambda}_{\mu\nu\rho} = \partial_{\nu}\Gamma^{\lambda}_{\mu\rho} - \partial_{\rho}\Gamma^{\lambda}_{\mu\nu} + \Gamma^{\lambda}_{\sigma\rho}\Gamma^{\sigma}_{\mu\nu} - \Gamma^{\lambda}_{\sigma\nu}\Gamma^{\sigma}_{\mu\rho} \tag{29}$$

While the coupling of matter fields to the gravitational variables is unambiguously determined, one can examine various possible choices for the gravitational action. One is free to set the torsion equal to zero from the beginning as a kinematical condition; then we have left the minimum extension of conventional general relativity needed to handle spinorial matter fields. If the torsion is present, its source is the spin density of matter; and whether or not it propagates depends very much on the choice of gravitational Lagrangian.

There have been generalizations of all this in several directions[28]. For instance one can put the entire Poincaré group, rather than just its homogeneous part, into the fibre; then the vierbein and the spin-connection can be looked upon to some extent as similar geometric objects, as the former become translational gauge potentials. Naturally we do not go into details here, but here we have another view of the summit.

General Relativity and Quantum Theory

Now I come to the last view of the summit that I wish to present. This has to do with an interesting comparison between quantum theory and general relativity.

As you may know, Bohr tried on many occasions to win Einstein's approval and support for his Principle of Complementarity, a key ingredient of his interpretation of quantum mechanics; but these of course Einstein never gave. That Bohr tried very hard to convince Einstein is clear from a phrase in his 1949 summing up of their years of debate when he says[29]: "The principal aim, however, of these considerations, which were not least inspired by the hope of influencing Einstein's attitude, was . . .". It is then interesting to see what attitude Bohr took towards general relativity! I began by saying that it is common to regard general relativity as the summit of classical physics. Bohr however viewed it differently. He felt that "The abstract character of the formalisms concerned is indeed, on closer examination, as typical of relativity theory as it is of quantum mechanics, and it is in this respect purely a matter of tradition if the former theory is considered as a completion of classical physics rather than as a first fundamental step in the thorough going revision of our conceptual means of comparing observations, which the modern development of physics has forced upon us"[30]. Presumably in the hope of convincing Einstein(!), he went even further and tried to draw analogies between the two theories. For instance he pointed out that in quantum mechanics, while object and apparatus are both ultimately quantum mechanical in nature, the theory of measurement makes us distinguish them and insist that the latter be describable in the limiting classical language; analogously in relativity while general covariance tends to obliterate the difference between time and space coordinates, we are ultimately obliged to recognise the physical distinction between them. As another instance he said that in relativity our description of any phenomenon depends in an essential way on the spacetime coordinate system, even though it may be supported by definite rules of transformation; the analogy in quantum mechanics is that complementarity limits our attempts to interpret results of experiments independently of the experimental apparatus or arrangement. In case it is just a bit difficult to follow this, I offer you two statements, one by Dirac on relativity and the other by Heisenberg on quantum mechanics, which may help. Speaking of the metric tensor components Dirac says[31]: "They determine both the coordinate system and the curvature of the space... . They describe not only the gravitational field, but also the system of coordinates. The gravitational field and the system of coordinates are inextricably mixed up in the Einstein theory, and one cannot describe the one without the other". Compare this with Heisenberg's statement on the nature of the wave function in quantum mechanics[32]: "This probability function represents a mixture of two things, partly a fact and partly our knowledge of a fact". There is some similarity in these statements , and this is possibly what Bohr was hinting at.

Conclusion

I would now like to conclude, in a manner befitting this place and this occasion, this series of long-distance views of general relativity I have tried to bring before your eyes. But here I am a victim of the phenomenon of "anticipatory plagiarism" which is described in a recent article by Gunther Stent in this way :[33] "Anticipatory plagiarism occurs when someone steals your original idea and publishes it a hundred years before you were born". It so happens that in a lecture to the Indian Academy of Sciences in 1985 S. Chandrasekhar already said exactly what I would like to say now. As there is hardly any chance of improving upon his expression let me quote him verbatim[34]: "The pursuit of science has often been compared to the scaling of mountains, high and not so high. But who amongst us can hope, even in imagination, to scale the Everest and reach its summit when the sky is blue and the air is still, and in the stillness of the air survey the entire Himalayan range in the dazzling white of the snow stretching to infinity? None of us can hope for a comparable vision of nature and of the universe around us. But there is nothing mean or lowly in standing in the valley below and awaiting the sun to rise over Kanchenjunga".

References

1. A. Pais, *Subtle is the Lord... The Science and the Life of Albert Einstein*, Oxford University Press (1982), Chapter IV, especially p.250 ff.
2. Ref.(1), p.257 ff.
3. See, for instance, W.Moore, *Schrodinger - life and thought*, Cambridge University Press (1989) p. 211, 221.
4. L. Rosenfeld, Ann. Physik **5** 113 (1930) ; Ann. Inst. H. Poincaré **2** 25 (1932).
5. P.A.M. Dirac, Proc. Camb. Phil. Soc. **29** 389 (1933).
6. P.A.M. Dirac, Can. Jour. Maths. **2** 129 (1950).
7. J.L.Anderson and P.G.Bergmann, Phys. Rev. **83** 1018 (1951).
8. P.A.M. Dirac, Proc. Roy. Soc. **A246** 333 (1958).
9. R. Arnowitt, S. Deser and C.W. Misner, Phys. Rev. **116**, 1322 (1959), **117** 1595 (1960).
10. P.A.M. Dirac, Scientific American **208** 45 (1963).
11. T.Regge, *The Group Manifold Approach to Unified Gravity*, Les Houches Lectures (1983).
12. P.A.M. Dirac, V.A. Fock, B. Podolsky, Phys. Z. Sowjetunion **2** 468 (1932); reprinted in *Quantum Electrodynamics*, J.Schwinger (ed.), Dover Publications, Inc., New York (1958).
13. See, for instance, *Quantum Electrodynamics*, J.Schwinger (ed.), Dover Publications, Inc., New York (1958).
14. P.A.M. Dirac, Revs. Mod. Phys. **21** 392 (1949).
15. H. Weyl, *Space Time Matter*, 4th edition (1922), reprinted by Dover Publications, Inc., New York (1952), § 34, 35; see also W.Pauli, *Theory of Relativity*, Pergamon Press (1958), p.192 ff.

16. V.A. Fock, Zeit. f. Phys. **39** 226 (1927).

17. F. London, Zeit. f. Phys. **42** 375 (1927); Naturwiss. **15** 187 (1927).

18. H.Weyl, Zeit. f. Phys. **56** 330 (1929).

19. C.N.Yang, *Geometry and Physics*, in *To Fulfill a Vision: Jerusalem Einstein Centennial Symposium on Gauge Theories and Unification of Physical Froces*, Y.Neeman (ed.), Addison - Wesley (1981).

20. O. Klein, in New Theories in Physics (International Institute of Intellectual Cooperation, League of Nations) (1938), pp. 77-93.

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

22. See, for instance, A. Trautman, *Fibre Bundles Associated with Space-Time*, Reports on Mathematical Physics 1 29 (1970); *Differential Geometry for Physicists*, Bibliopolis, Napoli (1984).

23. E. Cartan, *On Manifolds with an Affine Connection and the Theory of General Relativity*, translated by A. Ashtekar, Bibliopolis, Napoli (1986).

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

25. T.W.B. Kibble, Jour. Math. Phys. **2** 212 (1961).

26. D.W.Sciama in *Recent Trends in General Relativity*, Pergamon Press (1962).

27. *Elie Cartan - Albert Einstein - Letters on Absolute Parallelism 1929-1932*, Robert Debever (ed.), Princeton University Press (1979).

28. For more details see, for instance, A.R. Prasanna, *Differential Forms and Einstein-Cartan Theory* in *Gravitation, Gauge Theories and the Early Universe*, B.R. Iyer, N.Mukunda and C.V. Vishveshwara (eds.), Kluwer Academic Publishers (Dordrecht) (1989).

29. N.Bohr, *Discussion with Einstein on Epistemological Problems in Atomic Physics* in *Albert Einstein: Philosopher Scientist*, The Library of Living Philosophers, Inc., Evanston, Illinois, Vol.7 (1949) p.199; reprinted in N.Bohr, *Atomic*

Physics and Human Knowledge, Science Editions, Inc., New York (1961), p.52.

30. N.Bohr, *Atomic Physics and Human Knowledge*, Science Editions, Inc., New York (1961) p.65.
31. P.A.M. Dirac, *General Theory of Relativity*, John Wiley, New York (1975), p.9, 26.
32. W.Heisenberg, *Physics and Philosophy*, Harper & Row, New York (1958), p.45.
33. G. Stent, *Complexity and Complementarity in the Phenomenon of Mind*, University of California (Berkeley) Preprint (1989); the phrase is due to Robert Merton.
34. S. Chandrasekhar, *Truth and Beauty - Aesthetics and Motivations in Science*, Chicago University Press (1987), p.26.