AN ELEMENTARY COURSE ON GENERAL RELATIVITY

Tullio Regge

Lectures given in the
Academic Training Programme of CERN
1982–1983

GENEVA
1983
© Copyright CERN, Genève, 1983

Propriété littéraire et scientifique réservée pour tous les pays du monde. Ce document ne peut être reproduit ou traduit en tout ou en partie sans l'autorisation écrite du Directeur général du CERN, titulaire du droit d'auteur. Dans les cas appropriés, et s'il s'agit d'utiliser le document à des fins non commerciales, cette autorisation sera volontiers accordée.

Le CERN ne revendique pas la propriété des inventions brevetables et dessins ou modèles susceptibles de dépôt qui pourraient être décrits dans le présent document; ceux-ci peuvent être librement utilisés par les instituts de recherche, les industriels et autres intéressés. Cependant, le CERN se réserve le droit de s'opposer à toute revendication qu'un usager pourrait faire de la propriété scientifique ou industrielle de toute invention et tout dessin ou modèle décrits dans le présent document.

CERN - Service d'Information scientifique - RD/602 - 4000 - Septembre 1983

Literary and scientific copyrights reserved in all countries of the world. This report, or any part of it, may not be reprinted or translated without written permission of the copyright holder, the Director-General of CERN. However, permission will be freely granted for appropriate non-commercial use.

If any patentable invention or registrable design is described in the report, CERN makes no claim to property rights in it but offers it for the free use of research institutions, manufacturers and others. CERN, however, may oppose any attempt by a user to claim any proprietary or patent rights in such inventions or designs as may be described in the present document.
ABSTRACT

This report gives an informal account of the theory of general relativity, for non-specialists. It does not contain any detailed technical exposé of tensor calculus but relies instead on a number of intuitive arguments.

After a brief historical introduction the notion of curvature is developed, first in two dimensions (as done originally by Gauss) and then in higher dimensions, following the ideas of Riemann. This curvature is then related to quantities of physical interest through the following steps:

i) The equality of gravitational and inertial masses is discussed and presented as the 'weak equivalence principle'.

ii) This is then extended to the 'strong equivalence principle' according to the original programme of Einstein.

iii) The 'strong equivalence principle' implies the existence of a local inertial observer in any point of space-time. In a sufficiently small region of space-time this observer will not sense any gravitational field.

iv) In a larger region the observer will, however, sense residual tidal forces. These forces are identified with the curvature of space-time, to achieve a direct geometrical interpretation of gravity.

v) Finally certain curvature components are related to the distribution of matter through Einstein's field equations.

Section 4 contains a discussion of the classical tests of the theory and of the possibility of detecting gravitational waves. Sections 5 and 6 deal with cosmology and with the possible extension of the theory along the lines of the original ideas of Einstein, with emphasis on the dimensional reduction techniques of current interest.
CONTENTS

1. HISTORICAL INTRODUCTION 1
2. THE EQUIVALENCE PRINCIPLE 2
3. THE FIELD EQUATIONS AND THE GRAVITATIONAL POTENTIAL 7
4. THE CLASSICAL TESTS FOR GENERAL RELATIVITY 13
5. COSMOLOGY 17
6. BEYOND RELATIVITY 24
1. **HISTORICAL INTRODUCTION**

The philosophers of science are still debating the issue as to whether special relativity began with the famous 1905 paper of Einstein or whether it existed before in the works of Lorentz and Poincaré. In fact, the notion of "corresponding states" contained in the 1904 paper of Lorentz in many ways anticipates relativity, while still clinging to the obsolescent notion of ether. However, there seems to be little disagreement with the assertion that Einstein was the almost single-handed creator of the theory of general relativity. But it can also be asserted that this theory has its roots in the far-reaching geometrical investigations of G.F.B. Riemann; in turn, Riemann was heavily inspired by the beautiful "Disquisitiones" of Gauss, a masterpiece, dealing with the differential geometry of curved surfaces. A central theme in the theory of general relativity is the notion that the presence of matter influences the geometry of space and that this cannot be considered anymore as Euclidean. If we look back we find that Einstein had predecessors who had strange and powerful hunches about what was to come. Riemann himself toyed briefly with the idea that real space was curved. The eminent physicist and physiologist H. Helmholtz (1821-1894) investigated the physical aspects of Riemann theory and put stringent limits, from astronomical evidence, on the curvature of space. The geometer W.K. Clifford (1845-1879), who invented Dirac algebras before Dirac, thought of matter as a sort of ripple on a curved space. Many of his ideas reappeared in general relativity. These attempts, no matter how brilliant, were obviously premature. Physicists lacked the idea of a space-time manifold and had not yet understood the central role of electrodynamics. A complete construction of a relativistic theory of gravitation was achieved only at the end of the First World War.

Einstein did not arrive easily at the final result and had to go through years of intellectual wanderings before writing down his field equations. Some of his dearest colleagues and friends even thought that he had "gone off", carried away by some crackpot fantasy. We can reasonably assume that he was interested in the equivalence principle as far back as 1911. When he returned from Prague to Zürich in 1912 he met Marcel Grossmann at the ETH and began studying Gaussian curvilinear coordinates and their generalizations. Through Grossmann he became acquainted with the algorithm of absolute differential calculus, developed by the Italian mathematicians Gregorio Ricci and Tullio Levi-Civita. In fact, it is known that Luigi Bianchi, a most influential figure in Italian mathematics at that time, was a thoroughly sceptical critic on the matter of absolute differential calculus, and that recognition of the work of Levi-Civita and Ricci came as a spin-off of relativity. After many unsatisfactory attempts the final version of the theory was eventually ready in 1916, one year after Karl Schwarzschild found the isotropic solution which bears his name and which replaced the Newtonian potential. Finally the theory got its most spectacular test during a 1919 expedition to Prince Island, with the participation of Eddington, confirming the deflection of the light rays around the sun during an eclipse. In the twenties the theory of general relativity was applied to cosmology. In his later years Einstein tried desperately to arrive at a unified theory of gravitation and electromagnetism. Although his work had a great philosophical and ideological impact on his contemporaries the attempt was clearly premature. The vast increase in our knowledge and novel theoretical ideas have started these efforts again, I hope, with better chances of finding a final spectacular synthesis.
2. **THE EQUIVALENCE PRINCIPLE**

It is customary to distinguish between a weak and a strong equivalence principle. The weak principle asserts the equality, modulo some trivial proportionality constant, of the so-called inertial and gravitational masses. The inertial mass \( m_1 \) is measured by attaching to a body a known force \( F \) and measuring its acceleration \( a \) from \( F = m_1 a \). Therefore \( m_1 \) measures the inertia of the body -- its reluctance to be set into motion. On the other hand, the gravitational mass \( m_g \) is deduced from Newton's law of gravitation. The force of attraction between two bodies (say the Earth of mass \( M \) and a stone of mass \( m_g \)) is given by

\[
F = \frac{G M m_g}{r^2},
\]

where \( G \) is a universal constant which is equal to \( 6.6732 \times 10^{-8} \) dyn\( \cdot \)cm\(^2\)/g\(^2\), or \( G/c^2 = 7.425 \times 10^{-29} \) cm\( \cdot \)g\(^{-1}\). Empirically we have \( m_1 = k m_g \), where \( k \) is a universal constant which we can set equal to one. As a consequence, the acceleration of gravity, \( g = G M/r^2 \), is independent of the body; and, not taking into account air viscosity, falling bodies will follow the same trajectories.

We are aware that all this was known to Galileo, and popular accounts of this much-romanced fact hinge on a triumphant stone-dropping party of Galileo, perched on the top of the leaning tower of Pisa. Modern historians of science are sceptical of this account; they rather think that Galileo deduced his equivalence principle from the behaviour of wooden balls rolling down sloping planes. However, experiments with the pendulum also carry as much information and much more conveniently. Later on Haygens and \( \Phi \)ewton also worried about the strange equality of inertial and gravitational mass. But modern investigations began with the Hungarian Baron \( \Phi \)ötvös, who established that the ratio \( m_1/m_g \) is one within the rather stringent errors of \( 10^{-8} \). The limit has now gone down to \( 10^{-12} \), thanks to the work of Dicke at Princeton. Quite possibly \( \Phi \)ötvös was rather optimistic; it is often alleged that the mere presence of his body in the laboratory would have introduced an error comparable to the final one he claims.

Einstein found this weak equivalence very striking and expanded it into a great principle of physics. His starting point was a gedankenexperiment involving an "elevator" (see Fig. 1). The elevator is really a finite laboratory endowed with a physicist, with meters, clocks, and all one needs to follow the behaviour of matter. The experiment is done in four stages. In the first, the elevator is placed in some region of space far away from any celestial body and drifts with uniform motion. The internal observer checks that free bodies move with uniform motion and have no acceleration. The second step is to take the elevator, place it in a gravitational field, and let it fall freely. Since the acceleration of gravity, as a consequence of the principle, is the same for all bodies, including the wall of the elevator, the internal observer has no way of distinguishing this set-up from the previous one. In the third stage the elevator is brought again into empty space and is uniformly accelerated by a rocket engine. All bodies inside will appear accelerated by an acceleration \( a \), which is exactly opposite to that of the elevator and is common to all bodies. If we now hang the elevator in a gravitational field with \( g = a \), the internal observer will again find motions in no way distinguishable from those of the third experiment. As we shall see, this result is by no means perfect and holds really
only for infinitesimally small elevators. It is nevertheless striking, and shows a certain degree of equivalence between inertia and gravity; these are equivalent if we look at the phenomenon from a purely local point of view. Strictly speaking the equivalence between the first elevator (a) and second (b), and between the third (c) and fourth (d) (see Fig. 1), holds only for small accelerations and speeds, involving non-relativistic mechanics alone. Einstein blew it up into a great scientific manifesto, assuming its unrestricted validity far beyond the original proof. He required the existence of a local reference system such that for any gravitational field the free fall of the elevator would transform it into an inertial system. Inside this elevator the laws of physics would have the same form as in special relativity for any field and for any form of matter. This assertion is the strong equivalence principle.

In order to make this assertion work we need to know something more on how we have to define space and time coordinates inside a falling elevator. For this reason I prefer to postpone a technical discussion until the idea of curvature has been introduced. In order to do this I have to go back to Euclid and his axioms. The celebrated 11th axiom asserts the existence of a parallel to a straight line through any given point. If we use it and draw a parallel through the vertex of a plane triangle to the opposite side we can prove the familiar and seemingly trivial theorem stating that the sum of the internal angles $\alpha + \beta + \gamma$ of the triangle is exactly $\pi$ (Fig. 2). The sphere and the pseudosphere of Lobachevsky share with the Euclidean plane many properties but not the 11th axiom, and the theorem must be modified. For simplicity I shall describe it at first on the sphere of
radius R. Here the plane triangle is replaced with a spherical triangle, whose sides are arcs of great circles or geodesics. The shortest path between two points on the sphere is always an arc of great circle; these arcs appear as the appropriate generalization of the straight lines and have many common properties. Now let $\alpha$, $\beta$, and $\gamma$ again be the internal angles of the generic spherical triangle; we have

$$\alpha + \beta + \gamma = \pi,$$

where $A$ is now the area of the triangle. Indeed an octant has $\alpha = \beta = \gamma = \pi/2$ and area $\pi R^2/2$. [On Earth such an octant is approximately provided by the North Pole, Quito in Ecuador, and Libreville in the Congo (see Fig. 3)]. If we keep the area fixed and let $R$ go to infinity we retrieve the plane result. We can write formula (2) as follows:

$$K = \frac{1}{R^2} = \frac{\alpha + \beta + \gamma - \pi}{A}.$$  \hspace{1cm} (3)

In this case Eq. (3) appears as a useful way to measure $1/R^2$ from the area and internal angles of a generic triangle on the sphere. So, although the triangle is chosen by us, the result has a universal meaning -- the radius of the sphere. This radius appears as a parameter describing the violation of Euclid's 11th axiom. On the pseudosphere we have to replace $1/R^2$ by $-1/R^2$ (Fig. 4). We may look at a generic surface and define the limit:

$$K(x) = \lim_{A \to 0} \frac{\alpha + \beta + \gamma - \pi}{A},$$  \hspace{1cm} (4)

where the angles and area now refer to a generic geodesic triangle on the surface. The quantity $K(x)$ is now a function of the place and is called the Gaussian curvature. It should be clear that the value of the curvature does not change if we bend the surface.
without tearing or stretching it, that is if we think of it as being constructed out of an infinitely pliable and thin material. The definition gives a value for $K$ which is independent of the imbedding of the surface and depends only on the intrinsic geometry of it. This explains the reason behind the great success of the idea. The intrinsic geometry of any (differentiable) surface is then locally describable by means of a single parameter $K$. Planes, cones, cylinders, and developable surfaces all have $K = 0$; they are locally the same surface. If we need to draw a map of a surface (for instance that of the Earth) we have a satisfactory result only if this has $K$ small; for a generic $K = 1/R^2$ we achieve a faithful map only if the region represented is much smaller than $R$ in extension.

Therefore, even in the presence of curvature we can limit ourselves to a small enough region where Euclidean geometry is approximately valid. We can transform Eqs. (3) and (4) in yet another useful way. Suppose we have a gun (well ... a vector) placed at the North Pole. We displace it along the surface of the Earth by keeping it "parallel" to itself but always tangent to the surface of the Earth, thus acting as two-dimensional ants knowing nothing of the external space. Here by parallel I mean that the angle between the gun and the side remains constant. If you try it on the North Pole - Quito - Libreville triangle you find that the gun returns at the North Pole turned by the angle $\pi/2$, that is again $\alpha + \beta + \gamma - \pi$. The result holds in general and gives rise to a different procedure, one which extends naturally to higher dimensions.

In $n$ dimensions we take an $n$-dimensional vector and carry it around a small closed loop of area $A$. It will return acted upon by some $n$-dimensional rotation or, if we think of Minkowsky space, by a Lorentz transformation. We may divide that change in the vector by the area of the loop and obtain a local curvature parameter generalizing $K$. This parameter will depend on the vector chosen and on the orientation of the loop. It must therefore contain additional indexing in order to achieve an appropriate "bookkeeping". Traditionally this local parameter is given the letter $R$, after Riemann who defined it first (this causes some confusion with the radius $R$). In general, the change of a vector is of the form:

$$\delta V^\mu = R^\mu_{\lambda,\sigma} A^{\sigma\rho} V^\lambda, \quad \mu,\lambda,\sigma,\rho = 1, \ldots, 4,$$  

(5)
where $\sigma$ and $\rho$ are indexes labelling the surface element $d\sigma^p$, $\mu$ and $\lambda$ label the generic component of the vector, and $R^{\mu}_{\lambda \sigma \rho}$ is the Riemann tensor for the n-dimensional space. In what follows I need only to state that there are actually just 20 independent components for the Riemann tensor in four dimensions. These are split into 10 short-range and 10 long-range ones (the reason will be explained later). Once they are known we know the intrinsic shape of the four-dimensional space. It is clear that this shape can be very complicated; the curvature can change not only from point to point but also depends on the orientation of the measuring loop. The Riemann tensor and of course $K$ are inverse areas and are measured in cm$^{-2}$.

From this value we can extract a "size" of the region in which we can confidently use Euclidean geometry. Let me put aside Riemann and his tensor temporarily and go back to Einstein and his elevator. Had Einstein lived to see Gagarin, and the men on the Moon, he would have used a spacecraft instead of an elevator. Indeed, inside a spacecraft orbiting Earth we obtain almost perfectly the conditions for an inertial observer as stated in the equivalence principle. We get them to be "almost" perfect but not absolutely. If the spacecraft is in a circular orbit the gravitational force must balance the centrifugal one:

$$\frac{GM}{r^2} = \omega^2 r$$.

This happens at the centre of mass of the spacecraft. Near the wall facing Earth, however, attraction is stronger but centrifugal force is weaker; the opposite is true on the other side. If the size of the craft is $\ell$ the residual acceleration is of the order of $3GM/r^3$. The coefficient $GM/r^3$ is measured in s$^{-2}$; therefore, if we divide it by $c^2$ we obtain a quantity which is an inverse area. Should the area of the craft approach this area this would mean that the residual accelerations are not so weak; they would accelerate a test body to the speed of light inside the spacecraft. So the area must be immensely larger than that of the spacecraft. In the case of an Earth orbit we find a value for the square root of this area of the order of 100,000,000 km, roughly the radius of the Earth's orbit around the Sun. It should be obvious that these residual accelerations in the field of the Sun are responsible for the tides of the Earth when considered as a spacecraft orbiting the Sun. It is striking that both tidal forces and curvatures are measured in the same unit, cm$^{-2}$, and that they both control in the same way the size of the flatness region (spacecraft) in which Euclidean geometry (inertial physics) holds. So we assume that indeed the residual tidal forces must be directly related to curvature. This curvature is very feeble in our environment and would have no or almost no effect on practical geometry. Yet the identification curvature = tidal forces is of a very deep significance. Here geometry and physics come into contact in a totally new and unanticipated manner. In accepting this point of view we must think of space as being gently curved; we can establish locally an inertial observer but never extend it to a global gravitational field, its size being dictated by the curvature.

Suppose we have two nearby spacecrafts in orbit around the Earth. If their distances from the Earth are not the same they will have different periods and they will be mutually accelerated. This means that we cannot really join the two crafts into a single one and still ask that there is a Lorentz transformation relating their inertial systems. The respective local coordinates and inertial systems are in general relatively accelerated, therefore non-linear in time and hence also in the space coordinates. In attempting a global description of the gravitational field we must be ready to accept non-linear
transformations of coordinates, i.e. generic sets of coordinates. This realization deeply struck Einstein, who struggled with it and its revolutionary consequences for a few agonizing years. In fact, if there is curvature there is no such thing as Cartesian or Minkowsky coordinates. How can we have straight coordinates on the smooth surface of a potato? The same happens in curved space-time. There is no royal choice of coordinates; therefore we must write the theory in such a way that it looks equally good in all bad systems of coordinates. This means that we must consider generic reference systems which are no longer physically realizable by means of inertial observers but which relate also to accelerated observers.

The appropriate formalism to deal with this situation is absolute differential calculus. I cannot give here a technical exposition of the formalism; lots of people seem to think that it is difficult; at the time of Einstein it was considered as the pinnacle of human abstract thinking. I find it just as appealing and just as difficult as FORTRAN. Similarly to FORTRAN, it creates addicts who love absolute derivatives. I would like, however, to describe briefly the main difficulty which has been solved by calculus. Physicists need to take derivatives of vectors. They do so by taking the change in the components and dividing it by the change in the coordinates. A "constant" vector field is one (they assume implicitly) that has constant components. This is not true anymore in generic coordinates. A trivial and yet illuminating case is that of polar coordinates. A vector having a constant radial component is not in general constant. Therefore in analysing the change in the value of the components of a vector we must take into account two sources for it. One is the true change in the vector; the other is the tilting of the local coordinates from point to point and the gradual change in scale of the coordinate grid. Calculus teaches us how this last contribution must be subtracted in order to retrieve the "true" variation. If there is Riemannian curvature present it is impossible to have a system of coordinates which is globally Euclidean. I would also like to recall that parallel transportation of a vector along a closed path induces an effective change in the vector; alternatively if we have to displace a vector from one point to another the result depends on the path chosen; parallel displacement is not integrable. What I said for a vector holds for a generic system; parallel displacement of a rod or any object yields the same object acted upon by a generic Lorentz transformation depending on the surface encircled by the loop (triangle) used in the displacement. There is no harm in doing this; the final state of the system is still physically possible if the initial one was possible. It should be repeated once more that these effects in the gravitational field of the Earth are very small indeed; they imply rotation angles and values of $\beta$ of the order of $\xi^2/R^2$, where $\xi$ is again the size of the spacecraft, $R = 10^8$ km.

As a final remark: let us notice that the estimated value for the Riemann tensor, $\mathcal{G}^\mu_\nu / c^2 r^3$, is really a set of second derivatives of the usual gravitational potential; as such they are not independent because of the Poisson equation in vacuo. Therefore there is a certain combination of the $R^\mu_\nu \lambda \sigma$ which is short range, i.e. it should be set proportional to the local density of matter, the "source" of gravity. Other components instead have a long-range behaviour and signal the existence of a gravitational field outside matter.

3. THE FIELD EQUATIONS AND THE GRAVITATIONAL POTENTIAL

From the previous discussion it is clear that there should be a relation between curvature and distribution of matter. Just as the behaviour of matter is influenced by the
properties of space, we also expect that the geometry of space should be determined by the distribution of matter. How closely it is determined is in fact still a matter of debate. According to Mach, who strongly influenced Einstein but who equally strongly disagreed with him on the matter of relativity, the inertia of a body should be a particular consequence of the gravitational field of all the other bodies in the universe. There is no absolute space (a view strongly propounded by Berkeley); there are only material bodies, and the empirical geometry which we use is just a convenient way of discussing the properties of material bodies. In fact, it can be shown that in the theory of general relativity the inertia of bodies depends to some extent on the distribution of all the masses in the universe. But it is also true that space does not disappear from the theory; indeed it plays the role of the gravitational field. The field equations which relate the Riemann tensor to the distribution of matter are of second order in the gravitational field (they appear in fact as far-reaching generalizations of the Poisson equation) and must be supplemented by appropriate boundary conditions.

These equations can be written as

\[ R_{\mu \nu} - \frac{1}{2} g_{\mu \nu} R = \frac{8 \pi G}{c^4} T_{\mu \nu} , \] (7)

where the contracted Riemann tensor \( R_{\mu \nu} \) (the short-range components) is equated with the energy-matter tensor \( T_{\mu \nu} \). If we forget about indices the above equation appears as

\[ R = \frac{G}{c^4} \rho , \]

where the curvature \( R \) is now measured in cm\(^{-2}\) and \( G/c^4 = 7.425 \times 10^{-29} \) cm/g, so that \( \rho \) is indeed the density of matter. Once we know the short-range components we solve the field equations, determine the shape of space, and compute the long-range components. These are of the order of \( \propto GM/c^2 r^3 \), where \( M \) is the mass of the gravitating body and \( r \) the distance from it, as exemplified by Eq. (6). In general, the tensor \( T_{\mu \nu} \) will have 10 components describing how matter moves and giving more information than the conventional Newtonian density. From this point of view, ordinary gravity appears as the analogue of electrostatics, and general relativity as the analogue of the full Maxwell theory. If we could observe celestial bodies moving with a speed comparable with that of light, then we would see new forces playing the role of the Lorentz force and "magnetic" components of the gravitational field. However, we should not pursue this analogy too far. The photon is a neutral particle; therefore it cannot be the source of itself. Anything which carries energy and momentum is also the source of a gravitational field; therefore the quanta of this field, the gravitons, must generate other gravitational fields. This means that there is a bootstrap mechanism and that the field equations in general relativity are highly non-linear and are very difficult to solve analytically.

Eddington had a beautiful way of representing the action and reaction of matter and space. He imagined space as being a sort of elastic sheet, tightly stretched and flat. We now put on this surface a heavy steel ball (the analogue of a star). The ball produces a depression on the sheet, deforming it. If we place another ball near the first, the two of them will tend to fall into the depressions which they produce. In this way the sheet will generate a long-range force between the two objects; this force is the analogue of the
gravitational force. A better understanding of the gravitational field requires a more detailed discussion of the metric properties of space.

Perhaps the simplest way to see this role is to consider the action principle in special relativity and to generalize it in order to include the gravitational field. In its elementary form the principle of minimal action for a free particle requires that the integral

$$ S = \int_{1}^{2} \frac{1}{2} m v^2 \, dt $$

should indeed be a minimum for the solutions of the equations of motion. In Minkowsky space the action (8) is replaced by (minus) the proper time interval

$$ S = -m c^2 \int_{1}^{2} dt \sqrt{1 - \beta^2} , $$

which admits Eq. (8) as a convenient approximation if we expand in $\beta^2$ and retain the first two terms. Of course, the leading term $-m c^2$ does not count since it is a constant. If the particle is not free we have of course in place of Eq. (8) ($\Phi$ is the gravitational potential):

$$ S = \int_{1}^{2} \left[ \frac{1}{2} m v^2 - m \Phi(x) \right] \, dt . $$

This raises the problem of how to generalize it to the relativistic case. One simple way to do it (and in fact the right one) is to sweep the potential $\Phi(x)$ under the radical and use it to define a variable speed of light:

$$ S = -m c^2 \int_{1}^{2} dt \sqrt{\frac{\beta^2}{c^2} - \beta^2} . $$

We remember now that $dt^2 (1 - \beta^2)$ really stands for the invariant proper time interval of Minkowsky space:

$$ c^2 \, dt^2 = c^2 (dt)^2 - (dx)^2 - (dy)^2 - (dz)^2 = c^2 \, dt^2 - (dr)^2 . $$
Therefore the action (11) looks as if it has a proper time interval of the form

\[ c^2 \, dt^2 = c^2 \left[ 1 + \frac{2\Phi(x)}{c^2} \right] (dt)^2 - (dr)^2 \]  

(13)

and is equally minimal \([\text{remember the minus sign in Eq. (9)}]\) for the solutions of the equation of motion.

One way of looking at expression (13) is to think of a variable speed of light

\[ c^2(x) = c^2 \left[ 1 + \frac{2\Phi(x)}{c^2} \right] , \]

(14)

and therefore of a refractive index

\[ n(x) = \frac{1}{\left[ 1 + \frac{2\Phi(x)}{c^2} \right]^{1/2}} . \]

(15)

This index is then higher in the valleys and smaller on the top of the mountains. Light travels faster in outer space than on Earth. If we changed the value of \( c \) everywhere by a factor of 2 nobody would notice; all clock standards would change by the same factor and the definition of the second would differ by a factor of 2, thus neutralizing any visible effect.

But we should certainly see an effect if \( c \) would be an effective function of the place. The factor \( (1 + 2\Phi/c^2) \) in Eq. (13) acts effectively as a variable refractive index. It also has a "metric" role similar to that of \( \sin \theta \) in the spherical coordinates on the sphere:

\[ (ds)^2 = (d\theta)^2 + (\sin \varphi)^2 (d\varphi)^2 . \]

(16)

In Eq. (16) \( ds \) is the distance between two points on the sphere, having coordinates \( \theta, \varphi \), and \( \theta + d\theta, \varphi + d\varphi \), respectively. It is quite clear that a given change of \( \varphi \) does not mean the same distance at the equator (where 1° is about 111 km) and at the poles (where it is zero). In order to get at the distance effectively travelled we must multiply by the conversion or "metric" factor \( \sin \theta \). In Eq. (13) the factor \( (1 + 2\Phi(x)/c^2) \) plays the same role; it converts changes in the time coordinate \( t \) into effective times elapsed. We cannot use the time elapsed as a coordinate, for its rate of change is not the same in all points and it is not possible to synchronize with \( t \) clocks running at different levels. Therefore, in a gravitational field we must use a universal time \( t \) in order to communicate between different observers. The variable speed of light is seen only if we use the time \( t \).

The principle of equivalence tells us that we see no effect if we work in a small enough region of space in a locally inertial system of reference. On Earth the refractive index varies only by an amount of the order of a thousandth of a millionth from unity and is equal to \( 1 - R/r \), where \( R = 2GM/c^2 \) is the so-called Schwarzschild radius of the mass \( M \). For Earth \( R \) is 0.88 cm, for the Sun about 3 km. A more physical interpretation for \( R/r \) is to see it as the \( R^2 \) of the escape velocity. If this approaches the speed of light then the refractive index accordingly becomes infinite and we have a total reflection. In this configuration it becomes more and more difficult for light to leave the body and a fortiori the same holds for any material body. The marquis de Laplace had already theorized, almost 200 years ago, that a spherical body with the density of water would have been capable of
withholding light if its radius would have reached the value of about 100,000,000 km. This value corresponds indeed to that of a body with \( R/r = 1 \), a so-called black hole. Inside a body of this size the curvature, as given by \( R = Gm/c^2 \), would be of the order of \( 1/r^2 \), where \( r \) is the radius of the body. This implies very strong deviations from Euclidean geometry. The following gedankenexperiment is also very illuminating concerning the strange phenomenon of different time standards. We construct a \textit{perpetuum mobile} machine with a pipe forming a vertical loop (see Fig. 5). The pipe is filled with a fluid of atoms having two states; the excitation energy is \( E \). The fluid ascends through the left column and returns through the right one. Ascending atoms are in the ground state; descending ones are excited. Once they get to the bottom they emit a photon which is then reabsorbed once they reach the top. Therefore the descending fluid is heavier than the ascending one and we obtain a net gain in energy of \( Eg/hc^2 \), where \( g \) is the acceleration of gravity and \( h \) the height of the apparatus. However, the machine works, and this gain in energy per atom is obtained, only if the emitted photon has the same energy as the absorbed one. In fact, in a gravitational field a photon must lose energy by exactly the amount \( Eg/hc^2 \) in order to save the principle of conservation of energy. The ratio of the emitted versus absorbed frequency is then given by

\[
\frac{\nu_1}{\nu_2} = \frac{\nu_{\text{emit}}}{\nu_{\text{abs}}} = \frac{n_1}{n_2} = \frac{\left(1 + \frac{2gh}{c^2}\right)^{1/2}}{\left(1 + \frac{2gh}{c^2}\right)^{1/2}}, \quad n_1 > n_2
\]  

(17)
where \( v_1, \phi_1, n_1 \) (\( v_2, \phi_2, n_2 \)) are evaluated at the bottom (top). This means, however, that the periods must be given by

\[
T_{\text{emit}} = \frac{T_0}{n_1}, \quad T_{\text{abs}} = \frac{T_0}{n_2},
\]

(18)
i.e. the period \( T_0 \), as measured in the \( t \) coordinate, is the same for both observers. An external observer, however, sees a photon coming out with a frequency which is red-shifted; he has the impression that all phenomena on the surface of the Earth (or for that matter on any celestial body) are slowed down. Vice versa, observers on the surface of the body have the impression that the outer time is running faster. So far the effect on Earth mountains is of the order of a few nanoseconds a day. Although very small it can be measured by an ordinary atomic clock (one of the first tests has been carried out between the city of Torino and the Plateau Rosa on the Matterhorn) or by gamma-rays (as in the Rebka and Pound experiment). At the extreme limit, where the body becomes a black hole, the effect reaches an infinite factor; time is effectively frozen at the surface. If a star implodes as a supernova it is quite possible that the nucleus reaches such a high density as to trigger the collapse to a black hole. In this case, the gravitational force wins over all other forces which resist compression. The radius of the nucleus decreases and approaches very quickly (a fraction of a second) the value \( R = 2MG/c^2 \). Once it gets closer to this value the collapse is seen as slowed down by an external observer; at the limit the process is stopped at the last frame of the film. The black hole is therefore a static, or almost static, object only if seen by external observers, neglecting quantum effects. An observer sitting on the surface would instead witness the collapse as happening in a fraction of a second; at the end of this period the evolution of the external universe would appear as accelerated by an infinite factor and he would see the end of the universe. But it would also enter into a region of space-time in which new and strange phenomena would occur, such as the close encounter with a true singularity of the metric; the observer would see the effect of infinite tidal forces and be destroyed by them. Also he would have no possibility of sending back a report on life inside a black hole.

This nice picture is modified by the idea of Hawking that black holes can evaporate into a gas of photons and other particles through a quantum tunnelling effect. The idea is that vacuum is not really vacuous; it is instead filled with pairs of particles/antiparticles. The life span of a pair is dictated by the Heisenberg uncertainty principle; if the energy of the pair is \( E \) then it will live about \( h/E \). If a black hole happens to be near the pair it can attract one of the particles of the pair and gobble it in the time \( R/c \), where \( R = 2MG/c^2 \) is the radius of the black hole. If \( R/c < h/E \) then the process is possible and the production of particles from the vacuum becomes possible; in a similar way a nucleus of sufficiently large charge \( Z \) (\( Z > 137 \)) can extract an electron from the vacuum and emit a positron. We expect, therefore, that a black hole will emit particles of average energy \( E \approx h/c^2/2MG \). This means that black holes will behave as bodies of temperature \( T = E/k \) and will radiate with a surface brightness going as \( T^4 \), i.e. with the inverse fourth power of the mass. But now the radius of the black hole is proportional to the mass and the total power should go as the inverse square of the mass. This means that big holes radiate very little and small ones a lot and that the life of the object should go as \( 1/M^3 \). A black hole of mass \( 10^{15} \) g should last about 10 thousand million years, the age of the universe. If such a black hole was formed during the Big Bang then it should be dying right now. The last few
seconds of life are characterized by an immense power of emission and it is reasonable to ask if it is possible to detect such events with a gamma-ray telescope or other means. If seen they would be direct evidence for the existence of black holes and direct proof of a number of interesting conjectures. So far the only objects which are likely candidates for black holes are binary systems, such as Cygnus X-1, with an intense X-ray emission. The features of this emission are highly suggestive of an accretion mechanism in which material from a blue giant falls into a black hole, is compressed and heated, and emits X-rays.

The gravitational red-shift has been seen in a number of objects. On the Sun it amounts to a part in a million and it is difficult, but not impossible, to see because of the thermal broadening of spectral lines. The effect is much more evident on white dwarfs and it should approach 10% on a neutron star.

From this discussion we see that a gravitational potential behaves in many ways as a varying refractive index. We must not forget, however, that the concept of refractive index is a poor substitute for the metric tensor. In turn this depends on the choice of the coordinate and has a certain amount of convention built into it. The concept of curvature is a very subtle way of extracting from the variation of the metric tensor what really is there and dropping inessential conventions.

4. THE CLASSICAL TESTS FOR GENERAL RELATIVITY

We have already discussed one of the tests of general relativity, i.e. the gravitational red-shift, which had in fact a great historical importance in fostering acceptance of the theory. It is regarded, however, as less specific than those tests regarding the bending of light in the field of the Sun and the precession of the perihelion of Mercury. Both these tests depend on crucial features of the gravitational field of the Sun and on specific assumptions on the motion of relativistic bodies in a gravitational field. I would like to comment first on this last argument. The equivalence principle fixes almost unequivocally the motion of point particles inside a gravitational field. We know that in an inertial frame it must move in a straight line, i.e. without acceleration. But also we know that the proper-time elapsed must be a maximum, so that action is a minimum. Both these requirements are actually seen to coincide and determine the motion along lines which extend the concept of a straight line to a curved space, the geodesics.

A geodesic on a surface is the shortest line between two points; it is also the straightest; an example is provided by the great circles on a sphere. It is then natural to require that bodies move along a geodesic in space-time. If we do so we see that indeed the geodesic motion is very close to the Newtonian motion and that it takes great ingenuity to see the difference. Sometimes people are puzzled by the remark that after all the shortest way between where the Earth is now and where it will be six months hence goes straight through the Sun; and why does it happen that instead Earth follows a nice elliptic orbit. Of course the ellipse is just the projection on ordinary space of a four-dimensional curve in space-time, which looks like a very elongated helix -- almost straight, in fact straighter than any other line. So the idea of "straight" applies to the four-dimensional picture and not to ordinary space alone. But let us come back to the motion of bodies. The great theoretical astronomer Le Verrier became famous for having predicted the existence of a new
planet, Neptune, from its perturbations on Uranus. As soon as observations revealed irregularities in the motion of Mercury he and others decided that there had to exist a new planet, Vulcan, the innermost one, producing the perturbations. Le Verrier even went so far as to predict from previous bits of information (possibly just sun spots) the actual orbit of Vulcan, and also announced that it would pass in front of the Sun at a given date. Nothing showed up and astronomers were disappointed and worried. In fact, the perihelion of Mercury precesses by the amount of 5600" per century, all of which but 42" are explained by the Newtonian perturbation of the other planets. If we compute exactly, using the equation of the geodesics and the Schwarzschild solution, the motion of the planets, we see that indeed the modified equations account for the extra 42" and predict similar effects for the other planets*), also in agreement with the observation. This agreement is very important since it depends on features of general relativity which do not exist in the Newtonian theory.

There exists an extension of general relativity, the so-called Iordan-Brans-Dicke (IBD) theory, where the effect is different and depends on a dimensionless constant appearing in the theory. If we want to accommodate this theory, we have to explain in part the precession of Mercury through other means. One way to do it is to assume some oblateness in the Sun; the resulting quadrupole term in the potential would produce a precession; the beauty of the test would be marred. Other observations in the solar system tend to exclude the IBD theory, and the direct measurements of the oblateness of the Sun are not conclusive. At the moment, the most economical explanation seems to be the one provided by the standard theory. Also the recently discovered binary pulsar, PSR 1916+15, consists most probably of two collapsed objects with masses of the order of 1.44 solar masses, moving along a highly eccentric orbit. The whole system could be contained inside our Sun. Under these conditions the precession reaches 4" per year and is quite easy to observe.

Albert Einstein finally became world famous after his successful prediction of the bending of the light rays grazing the limb of the Sun. In one of the early versions of the theory he predicted a value which is actually only half the correct one (about 1.75" of an arc). Luckily the First World War prevented an early test, which would have discredited the theory. In 1919 Eddington finally went to Prince Island, in the southern hemisphere, and

*) The general formula is \( \Delta \alpha = 6\pi G M / (t^2 L) \) radians per century, where \( L = (1 - e^2)a \), \( e \) being the eccentricity and \( a \) the major semi-axis.
indeed his observation agreed with the (correct) value. It must be noted that if we compute
naively the bending as that of a non-relativistic bullet flying past the Newtonian field of
the Sun we find the halved value. The difference comes from extra terms in the metric tensor,
which multiply the differential of the space coordinates (the standard one multiplies dt).
These terms are usually ignored in the motion of slow bodies; they are, in fact, multiplied
by factors of the order of $\beta^2$. They are, of course, very important in the motion of light
particles; their appearance confirms the deviations from the naïve Newtonian theory.

Here follows, however, the naïve computation of the bending. A bullet having velocity $c$
follows an approximate straight line with impact parameter $r$. We assume that the motion
takes place in the $x,y$ plane and that the particle travels along the $x$ axis. The component
of the acceleration along the $y$ axis is then

$$a_y = -\frac{GM_{\odot}}{(x^2 + r^2)^{3/2}} y,$$

(19)

where $y = r$, $M_{\odot}$ is the mass of the Sun. If the bending is very small (as we expect) then we
can neglect the change of velocity along the $x$ axis. The total change of velocity along the
$y$ axis is then

$$\Delta v_y = -\int_{-\infty}^{+\infty} \frac{GM_{\odot}}{(x^2 + r^2)^{3/2}} r \ dt = -\frac{GM_{\odot} r}{c} \int_{-\infty}^{+\infty} \frac{dx}{(x^2 + r^2)^{3/2}} = -\frac{2GM_{\odot}}{cr}$$

(20)

and the final deviation is

$$\theta = -\frac{\Delta v_y}{c} = -\frac{2GM_{\odot}}{c^2 r} = \frac{R}{r} \approx 0.86\".$$

(21)

Here appears again the factor $R/r$, which controls all relativistic phenomena. As I said,
the actual bending is twice as much.

The measurement is done in two steps. One waits for a suitable total eclipse of the
Sun and takes a picture of the surrounding field of stars. The picture is then compared with
one of the same field taken six months later, when the effects of the Sun are minimal. The
results of the observations agree within about 10% with the predicted value. More recently
the test has been conducted on a couple of quasars which are periodically occulted by the
Sun. This is done by using a radiotelescope and one does not have to wait for an eclipse.
The observation is again in good agreement with the theory. The gravitational bending is,
of course, present wherever there are masses. There is general agreement on the fact that
some multiple quasars showing several components with absolutely identical red shift are
actually multiple images of the same object created by the gravitational lens effect of an
interposed galaxy. There is, on the average, an optical path difference of a few months
between different images of the same object. If we could store the phase information long
enough we could carry out a gigantic cosmic interference experiment. The gravitational
bending should create ultimately a cosmic blurring of all images of far-away objects.

Other tests on general relativity have been conducted within the solar system and
concern delays in the echo of radar impulses reflected by planets, transmission of signals
by space probes, and subtle effects on the motion of the moon. There is good agreement
with the theory and the accuracy of these observations is continually improving. This means that general relativity is here to stay with us. It is in fact already part of technology. The global system of positioning, which is being planned now, uses a set of orbiting atomic clocks beaming accurate time signals down to Earth. The position of any receiving station is then deduced from the respective time delays within a few metres. This implies, however, that the gravitational blue shift has to be taken into account seriously, or else there will be a resulting cumulative error.

I come finally to a different subject. Just as the Maxwell field has quanta associated with wave propagation, also the gravitational field is supposed to have wave modes and associated gravitons. The field equations for gravitation are non-linear and this makes it difficult to give exact solutions for arbitrary gravitational waves. If we limit ourselves to the more realistic case of weak waves then it is quite possible to solve the field equations in the linearized approximation; this is certainly adequate for most purposes. Gravitational waves travel with the speed of light and, just as light, have two distinct polarization states. The analogy stops here. Gravitons are particles of spin 2 and their behaviour under rotation is quite different from that of photons. This means that if we rotate the field along the line of propagation it comes back after 180°. For photons we need 360°, for electrons 720°; in general we need 360°/J, where the spin of the particle is J. There are also related differences in the mechanism of emission. These can be seen as follows. The electromagnetic power $P$ emitted by an oscillating electric dipole of moment $p$ is given by

$$ P = \frac{1}{3} |p|^2 \omega^5 \frac{1}{c^3}, $$

(22)

The corresponding formula for gravitational emission is, instead, of order $\omega^8$:

$$ P = \left(\frac{ML^2}{TT^3}\right)^2 \frac{1}{\rho_0}, $$

(23)

where now we deal, however, with a quadrupole emission. There is, in fact, no dipole radiation for gravitational waves and, of course, no monopole, as in the Maxwell field.

In Eq. (23) there appears the universal power $P_0 = c^4 / G \approx 10^{58}$ erg/s, a very large one in ordinary units. On the other hand, the term $ML^2/T^3$ can be understood as the third derivative of a moment of inertia and is the power transferred within the system as (internal kinetic energy)/(period of oscillation). This means that the power radiated is always extremely small. The radiated power becomes an appreciable fraction of $P_0$ if we move back and forth any mass $m$ at relativistic speed at the Schwarzschild frequency $c^3/mG$. Conditions remotely approaching this standard are rare indeed; perhaps a (non-isotropic) supernova collapse appears as a likely candidate for a potentially observable emitter of radiation. It is instructive to compute the gravitational output of known objects. The exact formula for a rod rotating with frequency $\omega$ is

$$ P = \frac{32\omega^5 I^2}{5c^3}. $$

(24)

This formula can be used qualitatively also for an orbiting system. In the case of the Earth it gives the ridiculous value of about 0.2 kW. This means that a gravitational collapse of the
Earth's orbit will take place in about $10^{24}$ years. Other causes, such as perturbations by the planets, are likely to act much before, say in $10^{10}$ years. In the case of the binary pulsar PSR 1913+16 the period of revolution is a few hours and the system would decay in about $10^8$ years. This brings it within the bounds of observation and indeed there is such an effect in full agreement with the theory. We are dealing, however, with indirect evidence for gravitational waves; we see the energy losses of the system emitting them but we would like to receive them in our laboratory. In order to do this many groups, stirred by the pioneering work of Weber, have started building gravitational receivers. These are typically big aluminium cylinders. The incoming waves (expected in the range of 1 kHz) will produce small oscillations in the shape of the cylinder which, hopefully, we can detect. The efficiency of the system is controlled by the branching ratio

$$\eta = \frac{\text{gravitational decay rate}}{\text{all other decays}}.$$ 

In a typical metal cylinder this ratio is about $10^{-74}$, as one can evaluate again from expression (23), and this explains the horrendous difficulties encountered in setting up a reasonable programme for gravitational astronomy. The ratio can be improved by cooling the cylinder or using superconducting cavities. However, we need to be very lucky to see a realistic event at the present time.

5. **Cosmology**

It is customary to start a lecture on cosmology by quoting the so-called Olbers' paradox. Actually, as Weinberg puts it, the paradox was first stated by de Cheseaux at the time of the French Revolution, and then rediscovered by Olbers in 1823. The paradox is very simple and yet very deep; it excludes immediately the naive cosmology which postulated an infinite universe, Euclidean and eternal in duration. Briefly stated, if we suppose that we live in this kind of universe and we imagine it to be filled uniformly by stars with the same average luminosity, then each spherical shell between $r$ and $r+dr$ would send us the same amount of light. Indeed the volume of the shell grows like $r^2$ and the apparent luminosity of a star decreases like $1/r^2$. But there are infinitely many shells and the integrated power received on Earth would be infinite. Actually the situation is not so disastrous; what the calculation really means is that in whatever direction we look we should see finally the surface of a star; the sky would appear uniformly bright as the surface, say, of the Sun. But we know that nights are dark. Interposing dark clouds of interstellar dust does not save us; very soon the dust would get heated and radiate as the stars around it. So the universe must have existed only for a finite amount of time, or stars fade with the distance quicker than we believe, or something else again occurs.

Big-Bang cosmology provides us with many ways of escape. But, before proceeding, let me remark that progress in cosmology follows from progress in astrophysics; without a realistic theory of star evolution we would not understand galactic structure. Without a clear idea of what a galaxy is, there is no modern cosmology. For this reason I always like to quote the historical debate between Curtis and Shapley on the nature of spirals, which developed in the twenties. Shapley believed that there was only one island universe, our galaxy, and had the idea that all spirals were actually objects within our galaxy. Curtis (but Kant had this idea before him) thought of them as island universes of their own.
I see this debate as the last chapter of the Copernican revolution, the last fight against geocentrism disguised as galactocentrism. The construction of the big telescopes and the work of Hubble finally settled the issue. Spirals are indeed other galaxies and the universe is filled with a dust of galaxies. These, and not the stars, are the fundamental grains of matter with which we have to deal when we think big. There are about ten thousands of millions of potentially observable galaxies. On the scale of 100 megaparsec their distribution is approximately uniform.

In a fundamental paper "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie published in 1917, Albert Einstein ventured into the no man’s land where relativity begins to say something about the universe. At that time Einstein thought of a universe filled with stars; the evidence about galaxies was yet to come. His arguments had nevertheless a great impact on cosmology. He proposed a model of a universe endowed with the so-called cosmological principle. This universe is filled uniformly with matter. It is possible to define in it a cosmological time \( t \) in such a way that local properties of the universe, averaged over a sufficiently large domain, depend only on \( t \). In particular, the density and temperature of matter, the curvature of space, and the metric should be the same everywhere but functions of \( t \). Under these conditions the very complicated field equations of general relativity simplify enormously, become tractable, and it is possible to solve them for the simultaneous evolution of matter and geometry. In fact, if we are concerned with the not too early instants of the universe we do not need the field equations at all; it is well known that most equations follow from a seemingly trivial Newtonian approximation. Before we discuss these equations I feel that we should have some clear geometrical idea on the shape of a "homogeneous" universe. The concept involves space only; the presence of matter in our universe provides us with the possibility of defining a cosmological time and this sets this coordinate apart from the others. This is no violation of the principle of special relativity. The principle of equivalence implies that special relativity holds in any sufficiently small local inertial system, and this is still true. Curvature effects, globally, work to the extent that is is possible to define special coordinate systems anchored to the curvature components and therefore ultimately to the matter distribution.

A homogeneous space in two dimensions is the ordinary sphere, i.e. the locus of all points of three-dimensional space whose distance from the centre is constant. Another such space is the plane. The plane can be conceived as a limit of the sphere where the radius becomes very large. There is another class of homogeneous two-dimensional spaces, the pseudospheres, discovered by the founding fathers of the non-Euclidean geometry, Gauss, Bolyai and Lobachevsky. It is impossible to construct the whole pseudosphere in the ordinary three-dimensional space; only relatively small pieces of it can be built. The best is to imagine it as the mass hyperboloid in Minkowsky space with one time and two spaces. It appears as an infinitely extended surface. All these constructions can be extended to several dimensions in an obvious manner. My comment hinges on the intuitive aspects of the geometry of these objects. The geometry on the pseudosphere can be obtained by changing in all formulae the (positive) curvature \( 1/R^2 \) of the sphere into the negative \(-1/R^2\) of the pseudosphere; the plane (or the Euclidean space) appears as the transition between these cases.

If we work in polar coordinates on the sphere there is another interesting property. As we all know, the circumference \( C \) of a circle is \( 2\pi r \), where \( r \) is the radius, i.e. the
distance from the centre. On a sphere of radius $R$ the corresponding formula is
\[ C = 2\pi R \sin \left( \frac{r}{R} \right), \]
where $r$ is now the length of the path "along the sphere" (not the chord). This implies that $C$ is smaller; by expanding we find
\[ C = 2\pi r - \frac{\pi}{3} \frac{1}{R^3} r^3 = 2\pi \left( 1 - \frac{K}{6} r^2 \right). \tag{25} \]
In fact on the pseudosphere the corresponding formula is
\[ C = 2\pi R \sinh \left( \frac{r}{R} \right) = 2\pi \left( 1 - \frac{K}{6} r^2 \right). \tag{26} \]
As the radius grows we see that $C$ grows exponentially if $K < 0$, while it has a maximum and then returns to 0 if $K > 0$. Spaces of positive curvature are compact (finite in area or volume), while those of zero or negative curvature are infinite in extent. Curvature controls the amount of "neighbours" of a given point. Surfaces of negative curvature have the form of a saddle, which accommodates more neighbours (Fig. 4). Whatever the value of $K$, homogeneous spaces of dimension $n$ have a group of symmetry with $n(n+1)/2$ parameters; there are 3 (6) parameters in dimension 2(3). This number of parameters is the same as that which holds for the description of all rigid motions of a plane figure or of a solid body, respectively. We can take a body, displace it or rotate it; each of these operations contains three parameters.

In cosmology this means that if we live in a homogeneous space then space looks the same at every point, but also means that the view, on the average, is the same no matter in which direction we look from the given point. These spaces of maximum symmetry are then also isotropic. In analysing the metric of these spaces it is customary to think of them as originating from a three-sphere or three-pseudosphere of curvature $\pm 1$ through scaling by a time-dependent factor. It is also standard practice to use polar coordinates centred in some conventional "pole", usually the Earth. Also this coordinate is the dimensionless ratio $y = \sin (r/R)$ or $y = \sinh (r/R)$; $R_y$ is then the effective radius producing a given circumference $C$ for the circle. In terms of $y$ and of the usual polar coordinates the space metric looks as follows:
\[ (ds)^2 = R(t)^2 \left\{ \frac{(dy)^2}{1 - ky^2} + y^2 \left[ (d\theta)^2 + \sin^2 \theta (d\phi)^2 \right] \right\}, \quad k = \pm 1, \quad K = \frac{k}{R(t)^2} \tag{27} \]
where $k$ now appears in the radial part only. If we set $k = 0$ we obtain Euclidean space. The only evolution is now in the overall scaling factor $R(t)$. Homogeneous models do not have transitions from positive to negative curvature or vice versa. If $k > 0$ then the universe is called closed; otherwise it is open.

Using the equation for the geodesic we see that if a galaxy is initially in some point of coordinates $y$, $\theta$, $\phi$, and the derivatives $dy/dt$, $d\theta/dt$, $d\phi/dt$ vanish, then $y$, $\theta$, $\phi$ remain constant through time. This implies that time evolution in these models scales all distances of material objects in the same proportion as $R(t)$. And since $R(t)$ is an increasing function of $t$ we see that this scaling is an expansion: the one, we hope, discovered by Hubble and announced in 1929. This expansion is related to the observed cosmological red shifts. In a homogeneous universe we have to add the time part of the metric to the space part (27):
\[ (dt)^2 = (dt)^2 - R^2(t) \left\{ \frac{(dy)^2}{1 - ky^2} + y^2 \left[ (d\theta)^2 + r^2 \sin^2 \theta (d\phi)^2 \right] \right\}. \tag{28} \]
If we had $R(t) = \text{const}$ and $k = 0$ we would find again old Minkowsky space in polar coordinates. Suppose now that a galaxy in the distant past, placed at some distance $y$, emitted a light signal in the direction of Earth, at the time $t = t_1$; the signal is then received here at the time $t = t_2$. We suppose for simplicity that along the line of sight $d\delta = d\phi = 0$. We have then $dt = 0$ and, therefore,

$$\frac{1}{R(t)} \, dt = \frac{1}{\sqrt{1-ky^2}} \, dy. \quad (29)$$

This equation can be integrated and gives

$$\int_{t_1}^{t_2} \frac{dt}{R(t)} = \int_{y}^{0} \frac{dx}{\sqrt{1-kx^2}}. \quad (30)$$

A second signal is sent at $t = t_1 + T$ and received at $t = t_2 + T'$. We have, also

$$\int_{t_1+T}^{t_2+T'} \frac{dt}{R(t)} = \int_{y}^{0} \frac{dx}{\sqrt{1-kx^2}} = \int_{t_1}^{t_2} \frac{dt}{R(t)}. \quad (31)$$

By subtracting we find the relation:

$$\frac{T'}{R(t_2)} = \frac{T}{R(t_1)} \quad \text{or} \quad \frac{1}{T} = \frac{1}{T'} \frac{R(t_1)}{R(t_2)}. \quad (32)$$

This means that the original frequency of light is now scaled as $R(t_1)/R(t_2)$. At small distances we can expand this equation as

$$\frac{T'}{T} = \frac{v}{c} = (1+Z) = \frac{R(t_2)}{R(t_1)} = 1 + \frac{R'(t_1)}{R(t_1)} (t_2 - t_1) = 1 + H \frac{d}{c}, \quad (33)$$

which gives us Hubble's law of proportionality between red shift $Z$ and distance $d$. $H$ is Hubble's constant. It is not constant at all; it is a function of the cosmological time $t$. Of course, we would have to wait millions of years before seeing any changes in $H$ in any given galaxy. If we look at distant galaxies we see them as they were in the past and hopefully we can detect changes in $H$. The present value of $H$ is, according to Sandage, about 55 km/(s·Mpc); according to de Vaucouleurs it is about 100 km/(s·Mpc).

$H$ gives us a provisional estimate of the age of the universe. If $H$ was really a constant then this would mean that all galaxies were at the same point in the past, $H^{-1}$ years ago, i.e. about 10-20 thousands of millions of years ago. But galaxies attract each other; the value of $H$ was higher in the past and therefore the speed of expansion was also higher. All this implies that $H^{-1}$ is really only an upper limit on the age of the universe, unless one modifies the field equations, for instance by introducing the cosmological constant. Also if we insist in going too much into the past we arrive at the point when galaxies did not exist; they had not yet condensed from the clouds of hot gas filling the universe.

In this model we see that much of the observation is controlled and consists of information about the function $R(t)$. The deceleration parameter $q$ is defined as $q = -\ddot{R}/(R)^2$. As I
hinted, \( q \) can be measured, in principle, by looking at distant objects; in practice, the evolution of a galaxy during many billions of years is a very risky subject and \( q \) can be anything in the range of \( \sim 0.1 \) to 0.6. If \( q \) is very large then galaxies decelerate strongly; this means that in the past they went faster. Their red shift should be higher than what is expected from Hubble's law or, equivalently, galaxies with a high red shift should look brighter. How bright is debatable, and this is now the root of the present uncertainty on the value of \( q \).

We can relate \( q \) directly to the distribution of matter in the universe. I suppose that this matter has no pressure (for instance radiation is negligible) and has density \( \rho(t) \). I could then write Einstein's field equations but would need tensor calculus. So I prefer a more elementary method based on Newtonian gravity but which, strangely, works just as well. I take a small sphere (with respect to \( R \)) of radius \( r = R(t)y \). By classical Newtonian arguments, the force it exerts on a galaxy of mass \( m \), placed on the surface of the sphere, is the same as that of a point in the centre having the mass of the sphere. This implies the force

\[
ma = - \frac{GmM}{r^2} = - \frac{4\pi Gm \rho(t) r^3}{3r^2} = - \frac{4\pi Gm R(t)}{3} .
\]  

(34)

The acceleration is now computed from \( R(t) \). The final equation is then:

\[
a = \ddot{r} = \ddot{R}(t) y \quad \text{i.e.} \quad \ddot{R}(t) = - \frac{4\pi Gm R(t)}{3} .
\]  

(35)

Here \( r \) does not appear and the equation has a universal meaning for \( R(t) \). Equation (35) can be rewritten as

\[
2q = \frac{8\pi G}{10^2} = \frac{\rho}{\rho_{\text{crit}}},
\]  

(36)

where

\[
\rho_{\text{crit}} = \frac{3H^2}{8\pi G} = 1.1 \times 10^{-29} \left( \frac{H}{75 \text{ km/s/Mpc}} \right)^2 \frac{\text{g}}{\text{cm}^3} .
\]

As expected, the deceleration increases with the density of matter in the universe. Equation (35) gives one integral:

\[
(\dot{R})^2 + k \dot{r}^2 = \frac{8\pi G}{3} \rho R^2 .
\]  

(37)

This equation also follows directly from the field equations. From Eq. (37) we have

\[
2q - 1 = \frac{k \dot{r}^2}{R \dot{r}^2} .
\]  

(38)

This means that \( k \) has the sign of \( 2q - 1 \); \( q > \frac{1}{2} \) means closed universe; \( q < \frac{1}{2} \) means open universe. So again we see that \( q \) is a very crucial parameter in any discussion of the model.

The present value of \( q \), as derived from Eq. (36), has a lower bound of 0.014 deduced from the virial theorem applied to galaxies. This is too low to close the universe and is lower than current estimates of \( q \) from Hubble's law. Quite conceivably there is hidden matter in the form of intergalactic hydrogen or concentrated in invisible collapsed objects.
Also there would be mass in the form of a neutrino sea. All these mechanisms would lead to a substantial increase in \( \rho \) and close the universe.

Equations (35) and (37) hold only for dust, i.e. matter without pressure. In the early universe there was also radiation, and we must modify the field equations in order to take the pressure \( \rho \) into account. We find instead of Eq. (35):

\[
\frac{dR}{dt} = - \frac{4\pi G}{3} \left( \rho + \frac{3p}{c^2} \right) R.
\]  
(39)

Equation (37) remains the same. We have the equation of energy conservation for matter:

\[
\frac{d}{dR}(\rho R^3) = -3 \frac{p R^2}{c^2}.
\]  
(40)

For dust \( p = 0 \) and \( \rho = \text{const}/R^3 \). If \( p = \rho c^2/3 \), as in the case of radiation, then \( \rho = \text{const}/R^8 \).

At small \( R \) radiation dominates the evolution of the universe. The turning point occurs at about \( t = 10^8 \) years. If we deal with the matter-dominated era then Eq. (37) can be integrated exactly and gives the parametric solution, where \( H_0 \) and \( q_0 \) refer to the present era at \( t = t_0 \).

\[
R(t) = R_0 \frac{q_0}{2q_0 - 1} (1 - \cos \theta)
\]
(41)

\[
t = \frac{1}{H_0} q_0 (2q_0 - 1)^{-\frac{3}{2}} \left( \theta - \sin \theta \right).
\]

This is the equation of a cycloid. For small \( t \) we have

\[
t = q_0 (2q_0 - 1)^{-\frac{3}{2}} \frac{\theta^2}{6}
\]

\[
R = \left( \frac{3H_0 q_0 t}{2} \right)^{\frac{2}{3}}.
\]

The universe will achieve maximum dimensions at \( \theta = \pi/2 \) and then undergo contraction. The "big crunch" occurs at

\[
t = \frac{2\pi}{H_0 (2q - 1)^{\frac{1}{2}}} = 2t_m.
\]

If \( q_0 > \frac{1}{2} \) then \( t_m = \infty \) and we get into the open universe. If \( q_0 < \frac{1}{2} \) we have the formulae

\[
R(t) = R_0 \frac{q_0}{1 - 2q_0} \left( \cosh \psi - 1 \right)
\]
(42)

\[
t = \frac{1}{H_0} q_0 (1 - 2q_0)^{-\frac{3}{2}} \left( \sinh \psi - \psi \right),
\]

which have the same behaviour at \( t \) small but show indefinite expansion. If radiation is present then the small \( t \) behaviour has to be modified in \( R \approx \text{const} \sqrt{t} \). Under any conditions a first reasonable hypothesis is that the evolution of the universe proceeded with the minimum increase in entropy and that we can assume a quasi-equilibrium at any value of \( t \). A good estimate for the temperature is then \( T = 10^{18} \text{ K} \ (t/s)^{-\frac{1}{4}} \).
The universe began at $t = 0$ with infinite density and infinite temperature. This is just a reflection of our ignorance in dealing with the problem and there is great uncertainty as to whether we can really extrapolate the known laws of physics to such an extreme case. In any case, after a few seconds $R$ expands to about 1 light year and $T$ cools to $10^{18}$-$10^9$ K. Under these conditions neutrons fuse with protons, leaving a mixture of hydrogen and helium with traces of deuterium and other nuclei. There follows a long period of cooling. After a million years $T$ drops below 4000 K and atomic hydrogen recombines; also at about this time the energy density of radiation drops below that of matter. Matter is now transparent to radiation and photons begin to circulate around the universe. As predicted by Gamow, these photons are still visible today with a red shift $Z$ of about 1000, which brings them into the centimetre range ($T = 2.7$ K). They are called fossil radiation. The present energy density of this radiation is in fact about $4.40 \times 10^{-29}$ g/cm$^3$, while that of matter could be estimated to be about $10^{-29}$ g/cm$^3$. There is no doubt that our age is heavily matter dominated.

From the observation of the 2.7 K microwave background we infer that the early universe was very homogeneous. This brings us to some final reflections. The principle of equivalence states that two bodies cannot move with a relative velocity higher than $c$ if they are in the same local system. It says nothing about galaxies lying at opposite points in the universe or anyway at distances comparable with $R(t)$. Indeed the velocity of the antipodal point in the closed universe is at $v = \frac{v}{R}$ for small $t$ this quantity diverges like $1/\sqrt{t}$. This means that in the early universe the expansion velocity started at $v = \infty$ and $t = 0$, and decreased quickly afterwards. This implies total lack of communication among different points of the universe. Is it then that the universe is homogeneous? How does a point know that it must have the same density of matter as any other point? Some cosmologies postulate a different evolution in the very early times in order to allow for mixing and homogeneity. Also the picture of an exactly spherical universe is an overly-simplified one.

Matter is distributed in clumps of galaxies, and this shows in the curvature. Also within these clumps the expansion is slowed down. The picture of the cosmos as an inflating balloon must show local irregularities. And in fact the gravitational collapse of a star in a black hole is a sort of local anticipated "big crunch".

The cosmic expansion is to be seen, as a general trend, on the scale of intergalactic distances and not on the atomic, planetary, or even galactic scale. Phenomena within a galaxy are almost totally unaffected by the expansion.

As we look up at the sky we see more and more distant galaxies. We see them as they were in the past. Man does not observe a $t = \text{const}$ section of the universe. He gazes at a light cone with the tip on Earth. If we look at objects with increasing values of $Z$ they appear firstly as receding and placed at a distance of the order of $Zc/H$. As we increase $Z$ we see more and more deeply into the past when the radius of the universe was much smaller than it is now. Therefore, even if we increase $Z$ and the coordinate $y$ of the object, we ultimately reach smaller distances. At the limit of $Z = \infty$ we would see the very moment of the Big Bang. Unfortunately we are stopped at $Z = 1000$; at this value we see the fossil radiation, a pale reminder of the original Olbers' paradox. We need neutrinos of impossibly low energy to see anything beyond this point.
6. BEYOND RELATIVITY

General relativity was so successful that Einstein soon thought about going even further. He was mainly worried about unifying gravity and electromagnetism into a single field. None of his attempts lasted long enough, and only recently electromagnetic forces have been unified but not with gravity, rather with the weak interactions, according to the theory of Salam and Weinberg. Gravity theories can be roughly classified as follows:

i) Non-relativistic theories such as Newton's gravity theory and a host of other generalizations. This trend is now very obviously obsolete.

ii) Theories with a non-symmetric metric tensor. This means that the angle between vectors a and b is not the same as the angle between vectors b and a. The difference depends on the electromagnetic field. Einstein favoured this possibility in one of his last attempts. Now it has practically fallen into oblivion.

iii) Teleparallelism. This is a new approach in which parallel transportation of a vector does not rotate it but displaces it. Instead of curvature we have torsion. Just as curvature is "rotation per unit surface" torsion is "displacement per unit surface". It has followers, and was also proposed by Einstein. Particle physicists do not like it.

iv) Higher dimensional theories. Already in the thirties Kaluza and Klein proposed a five-dimensional theory of gravity, which served as a prototype for many theoretical papers and is currently undergoing a revival under the conventional name of "dimensional reduction". One assumes that besides the three space dimensions there is one (or perhaps two or even seven) more, so that space-time has effectively five or more dimensions. However, this fifth dimension is rather strange; otherwise it would not be distinguishable from the others. Kaluza and Klein suppose that space-time is the analogue of a cylinder with the four usual dimensions running along the height of the cylinder, the fifth along its circumference. The circumference of the circle, as it turns out, is ridiculously small -- of the order of Planck's length, i.e. \( L = \sqrt{\frac{G}{c^3}} \approx 10^{-33} \) cm. This explains why we cannot turn the wheel and go into the fifth dimension. This length is at the same time the Compton wavelength and the Schwarzschild length of the same mass \( M = \sqrt{\frac{G}{c^3}} \approx 10^{-5} \) g (the energy equivalent \( c^2\sqrt{GM/c^3} \) is roughly equal to the chemical energy of a tankful of petrol).

The motivations behind this choice of geometry are the following. A point particle has a component of the linear momentum along any dimension of space-time; the one along time is the familiar energy. If we live in five dimensions this means that there is a fifth component of the linear momentum telling us the state of motion of the particle along the extra dimension. The idea is that this component should be interpreted, modulo some universal constant, as the charge of the particle. The discreteness of the charge is then naturally tied up with the finite circle-like character of the fifth dimension. This is a quantum mechanical effect. Any quantum particle confined in a finite dimension acquires a discrete momentum along this direction, just as a piano chord vibrates only in a set of discrete modes, or just as an electron in the hydrogen atom has a finite number of wavelengths along any closed orbit. Once we assume that five-dimensional space is indeed a "cylinder", in the sense described above, we can introduce this description into Einstein equations in five dimensions. There will also be components of the gravitational field along the fifth dimension and these will
depend on the state of motion of particles along the fifth dimension, i.e. on their charge. This leads us to believe that these components are actually representing the electromagnetic field. Indeed a great formal success of Kaluza and Klein was to show that the action in five dimensions, if written separately in the usual four plus the extra one, neatly separates into the standard Einstein gravity and Maxwell action. The length of the fifth dimension is then known once we fix the strength of the electromagnetic interaction, and is related to Planck's length. The theory has enough interesting features to be worth developing. But we already know that is has many shortcomings. It predicts unreasonably high values for the charged excitation modes of the field (close to Planck's mass) and it needs extra dimensions to accommodate weak and other interactions. There is nothing particularly scandalous in admitting many extra dimensions; currently the trend is to have seven extra dimensions where many nice coincidences occur. The mathematical formalism is just as difficult as in the current four dimensions. And once we know that the world is very short along these dimensions we do not run the danger of confusing them with the others and ask why we did not see them before. At the very high energies of a tankful of petrol (on a single particle, however) we could explore phenomena occurring in the extremely small domains and gaze directly into the fifth dimension or more. The relevant energies are, however, well beyond the capabilities of present accelerators, including our galaxy.

Acknowledgements

I wish to thank Tatiana Fabergé and Georges Boixader for their beautiful illustrations.