Editor
JOSHI

Horizons of Physics

WILEY-

1

Horizons of Physics

Editor
A W JOSHI

SPONSORED BY

1

THE THEORIES OF RELATIVITY

JAYANT V. NARLIKAR
Tata Institute of Fundamental Research,
Bombay 400 005

I. The Special Theory
1. The inertial observers
2. Conflict with electromagnetic theory
3. The principle of relativity
4. Spacetime measurements
5. The four-dimensional spacetime
6. Relativistic mechanics
7. The present status of special relativity
II. General Relativity
8. Gravity and inertial observers
9. Non-Euclidean geometry and gravity
10. Einstein's equations
11. The Schwarzschild solution
12. Cosmology
III. Conclusion

The title of this chapter refers to 'theories' in plural because there are two theories of relativity, special and general, both proposed by Albert Einstein. The special theory proposed in 1905 is hailed as the greatest revolution in physics since the time of Isaac Newton. The general theory which followed a decade later is considered an intellectual exercise of the highest calibre. What were these two theories about? What is their present status in science? To what extent have they influenced the overall thinking in physics? We will take up these questions after describing the theories first.

I. The Special Theory

1. The Inertial Observers

To understand the motivation of the special theory of relativity let us

first consider two observers A and B in uniform relative motion with respect to each other. To fix ideas, we will adopt the Newtonian concepts of absolute space and absolute time to begin with. Accordingly we will take an origin O in space and choose rectangular coordinates (x, y, z) to specify the position of a typical point while t denotes the time which flows evenly as measured by all observers.

In our example above, let us simplify the problem by assuming that both A and B move in the x-direction with velocities v_A and v_B respectively. Suppose further that both were at O at the same time, which we take to be $t = 0$.

It follows that at any time $t > 0$ the x-coordinates of A and B are given by

$$x_A^0 = v_A t, \quad x_B^0 = v_B t. \tag{1}$$

Relative to A the displacement of B in the x-direction is given by

$$x_B^0 - x_A^0 = (v_B - v_A)\, t. \tag{2}$$

The quantity $(v_B - v_A)$ measure the relative velocity of B with respect to A in the x-direction.

Now, in practice we measures relative rather than absolute velocities. When we talk of a car moving at 50 kmph we imply motion relative to the earth. When we talk of the earth completing one orbit in the year, it is relative to the sun. The sun itself moves round the centre of our galaxy, and so on ....

Newton had realized the important fact that all measured motion is relative. His characterization of an absolute space was more an abstraction than reality; he could not suggest an experiment that would enable us to fix an absolute standard of rest. The reason for this difficulty is not due to inadequate technology. It arises because the basic laws of physics do not seem to depend on our knowing the absolute standard of rest—if it exists at all. Thus in our example *neither* A *nor* B *should be able to measure* v_A *or* v_B. They can of course measure their relative motion, that is, the difference $v_B - v_A$.

The above circumstance contains the germ of the principle of relativity: that the laws of physics are the same for A and B or for that matter, for any other observer in uniform relative motion with respect to them. It is convenient to identify a special name for such observers. These are the so called *inertial* observers: those on whom no external force acts. It is easy to see that by Newton's second law of motion these observers are *unaccelerated* relative to the absolute standard of rest; and so they are moving with uniform velocities with respect to one another.

So we may state the *principle of relativity* as the fact that the basic laws of physics are the same for *all* inertial observers.

Let us proceed further and consider a practical problem. Suppose observer A has found certain laws of physics which describe the changes

of certain measurable quantities with respect to space and time. Suppose he sets up a rectangular cartesian coordinate system with himself at the origin. The observer B does likewise. Again, to simplify matters let us suppose both these coordinate systems to be parallel to the original absolute system O_{xyz}. Then from equation (2) it follows that a point in space which at time t has coordinate x_A in A's reference frame will have coordinate

$$x_B = x_A - (v_B - v_A)\, t \tag{3}$$

in B's reference frame. The y, z coordinates will be the same in both reference frames.

Our next step in the argument is to drop references to absolute space (whose existence we cannot seem to detect) and state the above relation in purely relative terms. So, let $(v, 0, 0)$ denote the velocity of B relative to A, a quantity that can be measured. The (x_B, y_B, z_B) coordinates in B's frame are then related to the (x_A, y_A, z_A) coordinates in A's frame by the transformation law

$$x_B = x_A - vt, \qquad y_B = y_A, \qquad z_B = z_A,$$
$$t_B = t_A. \tag{4}$$

We have introduced the time equality in the two frames explicitly to emphasize the Newtonian view that time is the same for all observers.

The transformation (4) is called the *Galilean transformation* and the principle of relativity says that *physical laws are invariant with respect to Galilean transformations.*

2. Conflict with Electromagnetic Theory

The above principle seemed reasonable to physicists in the post-Newtonian era, until it ran into serious conflict with the laws of electromagnetism as propounded by Maxwell. These laws describe how the electric and magnetic fields **E**, **B** change with time and across space as a result of static and moving electric charges. Suppose one inertial observer A writes down Maxwell's equations in their standard form (in Gaussian units)

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad \nabla \cdot \mathbf{B} = 0,$$
$$\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0, \qquad \nabla \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{j}. \tag{5}$$

Here ρ is the electric charge density and $\mathbf{j}$ the current density in space, and c is the speed of light.

How will these equations look to another inertial observer B moving with velocity $\mathbf{v}$ relative to A ? The operator ∇ denotes spatial derivatives which will change as a result of the Galilean transformation (4). A little algebra will tell us that the equations as 'seen' by B in this rest frame will look different—and more messy—compared to (5). In particular, terms involving $\mathbf{v}$ will occur explicitly in B's set of equations.

Clearly not only is this aesthetically unpleasant, but it also raises the question: are we violating the principle of relativity? Why should B's equations be different from A's ? Moreover, if we solve (5) we discover that electromagnetic waves travel in A's reference frame with the speed of light c. This value will be different in B's frame and would depend on the direction of propagation.

This conflict suggests two possible ways of interpretation. The first and the obvious one to anybody accustomed to the Newtonian ideas of space and time is that, perhaps, the principle of relativity is wrong after all and that here we have a way of detecting the absolute rest frame. In the nineteenth century, the physicists looked for a mechanical explanation for any physical phenomenon. Was light itself of mechanical origin like sound? Sound waves propagate as compression-rarefaction waves in any compressible medium. What about light? Did it also require a medium? The mechanistic analogy led to the postulate of such a medium which was called *aether*.

So in the above example of two observers, one may be at rest relative to the aether and the other moving relative to it. Electromagnetic experiments should be able to detect such motion. If we do find such an effect then we may abandon our requirement of the principle of relativity.

The celebrated Michelson-Morley experiment was set up in 1887 to detect the earth's motion relative to the uniform aether background. In this experiment, shown in Figure 1, a beam of light was split into two parts, one travelling in the East-West direction and the other North-South. The two parts were made to recombine and interfere with each other to form interference fringes. Because the light velocity is different in different directions, the fringes were expected to shift when the whole apparatus was turned through a right angle.

No fringe shift was detected even though the apparatus was sensitive enough to find it. Thus, *prima-facie*, the first alternative was ruled out.

In 1887 this circumstance presented a puzzle and physicists like Fitzgerald and Lorentz tried to find an answer without seriously disrupting the Newtonian framework. However, all these solutions were of the patchwork kind. It needed Einstein's master-stroke to solve the riddle.

3. The Principle of Relativity

As it happened, the solution offered by Einstein was not inspired by the Michelson-Morley experiment, although, as we shall see later, the result of the experiment followed logically from it. Einstein was concerned more with the structure of Maxwell's equations and the problem we have posed earlier: how to resolve the apparent inconsistency between those equations and the principle of relativity.

As Einstein pointed out in his classic paper of 1905, the Maxwell equations can be made invariant between two inertial observers, provided

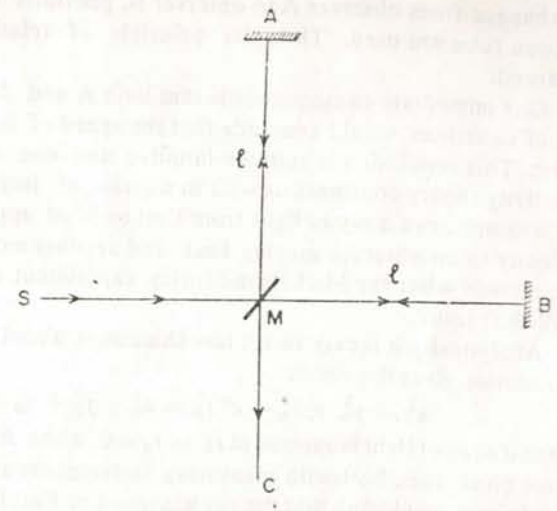

Fig. 1. In the Michelson-Morley experiment the light beam from source S is split at M where it is partially reflected along MA and partially transmitted along MB. Each part travels length l and is reflected at A and B. The reflected parts recombine at M and are observed along CM. The direction MA is along North and MB along the East. Using Newtonian mechanics and Galilean transformations we find that the two-way journey along the East-West route takes time $2lc/(c^2-v^2)$ while the two-way journey along the North-South takes time $2l/\sqrt{c^2-v^2}$. Here v is the speed of the earth in the West-East direction as it revolves about its axis in aether. Since $v \ll c$, the time difference between the arrival of the two components back at M is $\simeq lv^2/c^3$. When the apparatus is rotated through a right angle, the directionality of the arms is interchanged and the net time difference $2lv^2/c^3$ corresponds to a shift of 0.4 of the distance between the interference fringes of yellow light. The experimental apparatus could have detected even smaller shifts but it found no shift at all.

the space-time transformation is different from the Galilean one given in (4).

The new transformation law is:

$$x_B = \frac{x_A - vt_A}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad y_B = y_A, \qquad z_B = z_A,$$

$$t_B = \frac{t_A - \frac{vx_A}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{6}$$

It can be verified that the formal structure of Maxwell's equations is

unchanged from observer A to observer B, provided the above transformation rules are used. Thus the principle of relativity is completely restored.

One immediate consequence is that both A and B, solving the same set of equations, would conclude that the speed of light is the same for them. This conclusion is counter-intuitive and one of the first that the relativity theory confronts us with in a series of incredible results. How, for example, can a ray of light from East to West appear to have the same velocity to an observer moving East and another moving West? Yet this is precisely what the Michelson-Morley experiment also implied with its null-shift result.

Analytically it is easy to see how this comes about from (6). A simple calculation gives the result

$$x_A^2 + y_A^2 + z_A^2 - c^2 t_A^2 = x_B^2 + y_B^2 + z_B^2 - c^2 t_B^2. \qquad (7)$$

Thus if a ray of light is emitted at $t_A = t_B = 0$ when A and B were at the same place, then both with measuring instruments at rest in their respective frames, would find that the ray has speed c. For, if the light ray arrives at (x_A, y_A, z_A) at time t_A as measured in A's rest frame, then the left hand side of (7) is zero. And so the ray would arrive at (x_B, y_B, z_B) in B's rest frame at time t_B and he will measure the answer to be c, also.

What is so special about light that it should have the same speed for all inertial observers? Why don't other entities show similar invariance—for example, why not have invariance of sound speed? The answer to this query is that what is invariant and what is not is decided by the formal structure of the underlying equations of physics. Light speed emerges as invariant because of the structure of the underlying Maxwell equations. The formula for speed of sound does not show this relationship in terms of the underlying gas dynamic equations. Thus the role of the constant c in (6) and its identification with the speed of light are not decided in an ad-hoc manner but arise naturally from the study of symmetries of the underlying equations.

We next turn to the counter-intuitive notion of relativity of time as implied in (6). Notice that $t_A \neq t_B$: in other words, time does not have a unique absolute status. There is no universal simultaneity: the set of spatial points with t_A = constant is not the same set as that given by t_B = constant.

4. Spacetime Measurements

The constancy of speed of light in all inertial frames forces us to radically alter our concepts of measurements of space and time. A few examples will illustrate the new *modus operandi*.

(*i*) *Time measurement via signalling*—Imagine A and B, two inertial observers moving apart with relative speed v, having been together at an instant when both their clocks showed times $t_A = t_B = 0$. Figure 2 illustrates the situation. At any subsequent moment $t_A > 0$, A may send a light signal to B informing him the time (t_A) on his clock when the signal was sent out. Suppose B gets this signal at some time $t_B > 0$. It follows that if this experiment is performed at several such instants t_B will be found to be proportional to t_A, say, $t_B = \alpha t_A$.

Now the principle of relativity implies a symmetry between A and B; that is if their roles were reversed, the same factor α will operate. So a signal sent by B at t_B by his clock will be received by A at αt_B by his clock. Suppose A and B inform each other of their own times when they send out their signals.

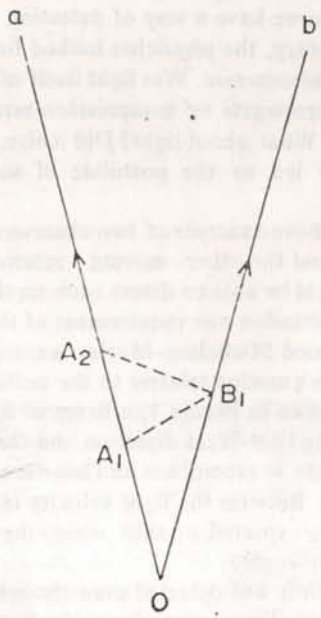

Fig. 2. The trajectories of A and B in spacetime are shown by two lines a, b intersecting at point O when $t_A = t_B = 0$. The signal sent by A at point A_1 is received by B at point B_1 and the signal reflected by B at B_1 arrives back at A at A_2. The tracks of light signals are shown by dotted lines,

Suppose A wishes to use signalling by light to measure B's speed. Accordingly he sends out a signal to B at $t_A = t_1$. B reflects it and the signal arrives back at A at $t_A = t_2$. A will then estimate that the signal had reached B at the half-way instant $(t_1 + t_2)/2$ and since light has travelled a total distance $c(t_2 - t_1)$, B's distance from A at the half-way stage was $c(t_2 - t_1)/2$. Thus A ascribes a velocity to B equal to

$$v = \frac{c\,(t_2 - t_1)}{t_2 + t_1}. \qquad (8)$$

However, by our earlier discussion, A's signal reached B at an instant αt by B's clock and B's reflected signal would reach A at

$$t_2 = \alpha(\alpha t_1) = \alpha^2 t_1, \tag{9}$$

by A's clock. Combining (8) with (9), we get

$$\alpha = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}. \tag{10}$$

This factor may be related to the well known Doppler effect. If A sends light waves of frequency ν, then the interval at the point of emission between two successive wavecrests is $1/\nu$. These crests are received by B at intervals α/ν by his clock. Therefore B will discover an increase in the wavelength of the arriving waves by the factor α. This is the relativistic Doppler effect. When $v \ll c$, we recover the Newtonian formula

$$\alpha \approx 1 + \frac{v}{c}. \tag{11}$$

It is also interesting for A to compare his estimate $(t_1 + t_2)/2$ of when his signal was reflected by B with B's own time αt_1 as communicated by him to A with the reflected signal:

$$\gamma \equiv \frac{(t_1 + t_2)/2}{\alpha t_1} = \frac{\alpha^2 + 1}{2\alpha} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} > 1. \tag{12}$$

In other words, A will conclude from this experiment that B's clock is running slower than his. A similar experiment conducted by B will lead B to the conclusion that A's clock is running slower than his! There is no contradiction between these conclusions since they are not drawn from one and the same experiment. Instead the argument reflects the inherent symmetry between A and B: neither has special status.

(*ii*) *The twin paradox*: The above observation of apparent slowing down of the clock of a moving observer viewed from the rest frame of another observer is called the phenomenon of *time dilatation*. The factor γ in (12) is the dilatation factor. It features in the transformation equations (6) and hence in many expressions involving measurements in relativity theory.

This factor comes in a well known paradox called the twin paradox. Suppose A and B are twins. A is an inertial observer while B moves away from A with uniform speed, travels a great distance, quickly reverses his speed and comes back with the same speed. If we ignore the short duration of the speed reversal, B's clock is running slow compared to A's and he will therefore have aged by $\gamma^{-1} T$ if A has aged by T. So B will be much younger than A if v is very nearly equal to c.

The paradox arises when we view the sequence of events in B's rest

frame. To him A appears to be moving away with uniform speed, quickly reversing his speed and returning back to the starting point. So by the same argument, should not A be younger than B?

Since A and B start and end at the same spacetime point the effect—whatever it may be—has to be an absolute one. So who is younger, A or B?

The paradox is based on the erroneous assumption that perfect symmetry exists between A and B. In fact, as mentioned earlier, A is an inertial observer and B is not. So the second part of the argument is not valid and B should be younger. Two questions are, however, raised by this paradox that take us beyond special relativity.

The first is, how does time-dilatation operate for a non-inertial observer? So far we have not discussed this point. As a matter of fact, one has to go to an extension of special theory to accelerated observers. This question really belongs to general relativity. Using such an extension one can in fact evaluate the effects on B's clock as seen by A, as B slows down and accelerates again. The conclusion that B is younger than A still stands.

The second, more fundamental question is how do we decide who is inertial? *If only* A *and* B *exist in the universe, there is no way we could answer this question*: for each one would legitimately find the other accelerated in his rest frame. It is only when we have a background of other objects that we may be able to settle this issue. We will return to this question towards the end while discussing cosmology.

(*iii*) *Spatial contraction* : Let us apply the transformation (6) to a rigid rod of length l at rest in the frame of B. A sees this rod move past with speed v in the x-direction. To fix ideas let one end of the rod be at the origin of B's coordinate system and the other have spatial coordinates $(l, 0, 0)$ in that frame. What length will A ascribe to the rod?

There are two ways of making this measurement. At a fixed instant, say t_A in A's frame, we note where the two ends of the rod will be. Thus in (6) we wish to know the values of x_A for $x_B = 0, l$ at the given t_A.

Thus we have for the two ends

$$0 = \frac{x_A^{(1)} - v t_A}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad l = \frac{x_A^{(2)} - v t_A}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

So,

$$l' = x_A^{(2)} - x_A^{(1)} = l\sqrt{1 - \frac{v^2}{c^2}} = l\gamma^{-1}. \tag{13}$$

Thus the rod appears to A shrunk by the factor γ^{-1}.

The second method involves noting the times $t_A^{(1)}$ and $t_A^{(2)}$ at which the front and the rear end of the rod pass a fixed point x_A in A's frame. The length of the rod is then deduced to be $l' = v\,(t_A^{(2)} - t_A^{(1)})$.

From (6) we have,

$$l = \frac{x_A - vt_A^{(1)}}{\sqrt{1-\frac{v^2}{c^2}}}, \quad 0 = \frac{x_A - vt_A^{(2)}}{\sqrt{1-\frac{v^2}{c^2}}},$$

i.e. $$l\sqrt{1-\frac{v^2}{c^2}} = v\,(t_A^{(2)} - t_A^{(1)}) = l'. \tag{14}$$

We get the same answer.

Notice that in the Michelson-Morley experiment the null result seemed to suggest that the path travelled by light in the E-W direction was shrunk by the same factor. In the pre-relativity days this result was thought to be absolute and was so proposed by Fitzgerald on empirical grounds. Why should a body contract by this factor in the direction of its motion? Lorentz offered an elaborate explanation for it in terms of electronic constitution of matter. Taking this contraction into account Lorentz in fact arrived at a modification of Galilean transformation, precisely that given by (6).

For historical reasons the transformation (6) is called the Lorentz transformation, although Einstein arrived at it from an entirely different angle. Thus the above 'contraction' as given by (13) or (14) is only relative: by reversing the roles of A and B the effect would still be there. That is, a rod at rest in A's frame will appear shrunk in B's frame. Although the contraction formula is the same the crucial difference between its interpretation in the Lorentz Fitzgerald and Einstein cases cannot be over-emphasized.

5. The Four-Dimensional Spacetime

The spacetime transformation law (6) is a special case of the more general situation where B is moving relative to A in arbitrary direction. We may write the spacetime transformation in the general case using the 4-dimensional notation of Minkowski. The trajectories of particles in spacetime are called *world lines* and points on it are called *world points*.

Let us use four coordinates x_A^l to indicate the time ($t_A = x_A^0/c$) and spatial position ($x_A^1 = x$, $x_A^2 = y$, $x_A^3 = z$) measured in the rest frame of world point A. The coordinates x_B^l for B's rest frame have similar interpretation. We have absorbed c in x^0 to make its dimension the same as that of x^μ ($\mu = 1, 2, 3$). In future we will use the Greek indices λ, μ, ν, ...etc. to label 1, 2, 3 while the Latin indices l, m, n, ...etc. will be used to label 0, 1, 2, 3. The relation (7) can be described by

$$\sum_l \sum_k \eta_{lk}\, x_A^l\, x_A^k = \sum_l \sum_k \eta_{lk}\, x_B^l\, x_B^k\,. \tag{15}$$

This relation is the 4-dimensional analogue of the famous Pythagoras's theorem. The 'metric' η_{lk} is a diagonal matrix:

$$\eta_{lk} = \begin{bmatrix} +1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix} \tag{16}$$

Note that we have chosen to attach positive sign to the time-like term and negative signs to the three spacelike terms, thus reversing all the signs in (7). It is convenient to introduce a 'summation convention' as follows.

For any quantity Q^l, define another with a lower index by

$$Q_l = \sum_k \eta_{lk}\, Q^k. \tag{17}$$

Now any summations of the kind

$$\sum_l P_l\, Q^l, \qquad \sum_k \sum_l R_{lkl} P^k Q^l, \ \ldots$$

may be abbreviated to

$$P_l\, Q^l, \qquad R_{ikl} P^k Q^l, \ \ldots \text{ etc.}$$

In other words drop the Σ symbol wherever an index is repeated above and below in a product expression just once. Thus the rule does not apply to sums like

$$\sum_l P_l\, Q_l, \qquad \sum_k P_{ik} Q^k\, R^k \text{ etc.}$$

With this shortening of way of writing the equality (15) becomes

$$x_A^l\, x_{lA} = x_B^l\, x_{lB}. \tag{18}$$

The generalization of (6) is the general Lorentz transformation

$$x_B^l = L_k^l\, x_A^k. \tag{19}$$

The coefficients L_k^l must satisfy a product rule in order to preserve (18):

$$\eta_{ik}\, L_m^l\, L_n^k = \eta_{mn}. \tag{20}$$

It is easy to see that (20) imposes a group structure on the elements (L_m^l). This group is called the Lorentz group. The spacetime in which this group operates and hence (18) is preserved is called the *Minkowski spacetime*. The principle of special relativity demands that all physical laws be invariant under the Lorentz group of transformations.

6. Relativistic Mechanics

Given the above requirement, we note that Maxwell's equations obey it; in fact their case motivated Einstein towards this general requirement. But what about Newton's laws of motion? These laws, as we saw earlier, were invariant under the Galilean transformation. They will have to be modified so as to be invariant under the Lorentz transformation. This requirement brings about radical changes in our thinking of the concepts of velocity, mass, energy, momentum, etc.

Rather than redefine these concepts under the formal structure of the general Lorentz group we will try to illustrate these with the help of the *special* Lorentz transformation of (6). Let us consider a particle at rest in B's frame, whose mass as measured in that frame is m_0. To A the particle appears to be moving with velocity v in the x-direction. What is its mass as measured by A?

Now in Newtonian parlance the velocity is a vector in 3-dimensional space. Here its components in A's frame are

$$v^a = \left[\frac{dx_A}{dt}, \frac{dy_A}{dt}, \frac{dz_A}{dt}\right]. \tag{21}$$

In the 4-dimensional space time the above concept is generalized to that of a 4-velocity

$$u^i = \left[c\frac{dt_A}{d\tau}, \frac{dx_A}{d\tau}, \frac{dy_A}{d\tau}, \frac{dz_A}{d\tau}\right] \tag{22}$$

Notice that the triplet (dx_A, dy_A, dz_A) is replaced by a quartet $[cdt_A, dx_A, dy_A, dz_A]$. But what is $d\tau$? We have already seen that time no longer has an absolute status in relativity, as it did in Newtonian mechanics. The next best that we can have in relativity is the *proper time* of the particle—the time it measures in its own rest frame. This is denoted here by τ. In our particular example τ is t_B.

So, using (6) we get

$$u^i = \gamma\,[c, v, 0, 0], \tag{23}$$

with γ given by (12). Just as the 3-velocity (21) transforms as a 3-vector under orthogonal spatial transformations, the 4-velocity (22) transforms as a 4-vector under Lorentz transformations. In particular if we restrict ourselves to special Lorentz transformation, we find (23) transforming as a vector.

(*i*) *Addition of velocities*: Let there be another observer C with respect to whom A is moving with velocity w in the x-direction. What is the velocity of the particle with respect to C? The Newtonian answer $v + w$ based on Galilean transformation is clearly wrong. To find the correct answer let us write the analogue of (6) between A and C;

$$x_A = \frac{x_C - wt_C}{\sqrt{1-\frac{w^2}{c^2}}}, \quad y_A = y_C, \quad z_A = z_C,$$

$$t_A = \frac{t_C - \frac{wx_C}{c^2}}{\sqrt{1-\frac{w^2}{c^2}}}. \tag{24}$$

Combining (6) with (24) we can write (t_B, x_B, y_B, z_B) in terms of (t_C, x_C, y_C, z_C). This gives us the answer. We find that the particle appears to move in C's frame with velocity in the x-direction given by

$$W = \frac{v+w}{1+\frac{vw}{c^2}}. \tag{25}$$

This addition-formula is consistent with the notion that light speed is c in all inertial frames. For, if $v = c$, then (25) gives $W = c$.

(*ii*) *Mass and energy* : The 4-momentum of the particle is defined as

$$p^i = m_0 u^i \tag{26}$$

In our special example,

$$p^i = m_0\,\gamma\,[c, v, 0, 0]. \tag{27}$$

Note that we 'recover' the Newtonian 3-momentum if we define the mass m in A's frame by

$$m = \gamma m_0. \tag{28}$$

Then the three space-like components of p^i are the same as the Newtonian momentum.

Using (27) for an accelerated particle also, we first note that along the trajectory,

$$p^i p_i = m_0^2\gamma^2\,(c^2 - v^2) = m_0^2 c^2 = \text{constant}. \tag{29}$$

Differentiating with respect to t_A, we get

$$0 = p_i \frac{dp^i}{dt_A} = p_0 \frac{dp^0}{dt_A} + p_1 \frac{dp^1}{dt_A},$$

i.e, $$v\frac{dp^1}{dt_A} = c\frac{dp^0}{dt_A}.$$

But if we use Newton's law of motion

$$\frac{dp^1}{dt_A} = F^1 \tag{30}$$

with F^1 as the impressed force, then we get for the above equation

$$v\,F^1 = \frac{d}{dt_A}(cp^0). \tag{31}$$

Now the left hand side of the equation is simply the rate at which work is being done by the force. The right hand side can then be identified with the rate of increase of energy of the particle. Thus it is natural to identify cp^0 with the energy E of the particle :

$$E = cp^0 = \gamma m_0 c^2 = mc^2. \tag{32}$$

This is Einstein's celebrated equation. For $v \ll c$, we get

$$E \approx m_0c^2 + \tfrac{1}{2} m_0v^2 = E_0 + T \qquad (33)$$

where $E_0 = m_0c^2$ = rest mass energy and $T = \frac{1}{2} m_0v^2$ is the kinetic energy.

We have arrived at these concepts in a heuristic fashion. A more rigorous derivation of these ideas can certainly be given. What we find in general is that ideas in mechanics have to be modified to take into account the relativity of time and space, and the validity of Lorentz rather than Galilean transformations. The Newtonian concepts of energy, momentum, force, etc are retained with modifications, the most significant of these being the mass-energy relation, Eq. (32).

7. The Present Status of Special Relativity

The special theory of relativity took some time to be assimilated into physics. In the early days, paradoxes like the twin paradox were evolved during discussions of the unusual aspects of the theory. These paradoxes have since been resolved although even today some tend to resurface under different garbs.

The assumptions of the theory and its consequences for mechanics and electrodynamics have been verified in high energy cosmic ray events as well as in accelerator experiments. Computations in high energy particle physics based on special relativity have been accurately verified and so no doubt exists today about the correctness of the theory.

Because the theory predicts significantly different effects compared to Newton's theory at $\gamma \gg 1$, and for macroscopic objects (e.g. humans travelling in spaceships) large energies are required to achieve these effects, the verification of the theory has been largely in microscopic areas. We end the discussion of the theory by noting that as $v \to c$, $\gamma \to \infty$ and $m \to \infty$. Therefore, infinite force is needed to accelerate a massive object to the speed of light. Hence the impossibility to achieve the light speed for any massive object.

There is an exception to this rule, namely, the photons—the particles of light: since they travel with the speed of light they cannot have nonzero rest mass. For them $m_0 = 0$. They do have energy and momentum, however. The relationship between energy E and momentum $\mathbf{p}$ for a photon is

$$E = |\mathbf{p}|\, c. \qquad (34)$$

The photon always travels with speed c and hence cannot be brought to rest. *The statement that the photon has a zero rest mass has therefore no operational meaning.*

Scientists have conjectured about the existence of a third class of particles, the *tachyons* which always travel with speed greater than c. These particles have not yet been detected, although their detection will not violate special relativity.

II. General Relativity

8. Gravity and Inertial Observers

Let us now turn to Einstein's real *magnum opus*, the general theory of relativity. At a superficial level the general theory may be said to achieve two things. It generalizes the notion of invariance of physical laws for *all inertial* observers to the notion of invariance of these laws for *all* observers. It also incorporates the theory of gravity in the process. At a deeper level, however, it is necessary to probe deeper for the motivation for this theory.

For, one might argue, why alter or generalize the elegant structure of special relativity ? Why alter Newtonian gravity when it has worked so well? The answer to these questions is that both these achievements are mutually incompatible.

Take special relativity first. It revolves round the so called inertial observers. Do such observers exist in reality? They don't; because gravity is an all-pervading and everlasting force that acts on every material object. At best the notion of an inertial observer is an idealization.

Second, Newton's law of gravity, either in its original inverse square form or as Poisson's equation, has the interaction propagating instantaneously. The law, in other words, is not Lorentz-invariant as would be required by special relativity.

Einstein felt that radical rethinking rather than patchwork modification was needed to resolve these conundrums. It indeed was a stroke of genius that led him to his solution, via spacetime geometry.

9. Non-Euclidean Geometry and Gravity

Basically, Einstein's argument could be rephrased thus. Granted that gravity is omnipresent and everlasting, it must be related to something intrinsic to spacetime itself since spacetime also shares the same two properties of gravity. What intrinsic feature of spacetime is related to gravity? Geometry! That was the solution offered by Einstein.

Geometry describes the rules of operations on spacetime measurements. The measurements of lengths, time intervals and angles, the notions of straight line paths, parallelism, curvature, etc. are part of geometry. Although the first and the simplest geometry to evolve was that of Euclid, there are other geometries as well. They all go under the comprehensive title of 'non-Euclidean geometries' although each geometry of this set can be quantified in as much detail as Euclid's geometry has been. It was Einstein's hunch that the phenomenon of gravity is linked with non-Euclidean geometry for spacetime. The following argument will illustrate the idea.

Suppose a projectile is thrown vertically upwards with speed v. At any time t measured since projection, the vertical height of the projectile is given by the parabola (see Figure 3)

$$y = vt - \tfrac{1}{2}gt^2. \tag{35}$$

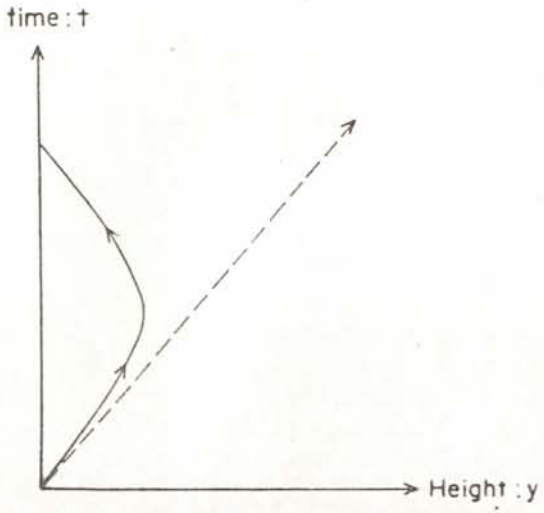

Fig. 3. The dotted straight line is the trajectory the projectile would have in the absence of gravity. The continuous line shows the actual parabolic trajectory.

This result is based on Euclid's geometry and Newtonian physics. If there were no gravity, Newton's first law of motion would have led to the trajectory

$$y = vt, \tag{36}$$

that is, uniform motion in a straight line. With gravity present, the projectile experiences downward acceleration g and so the trajectory in (t, y) plane is a parabola as given by (35).

Einstein would argue that since gravity cannot be eliminated, the straight trajectory (36) has no real status. The only relevant trajectory is that given by (35). Can we consider it as a 'straight' trajectory in (t, y) plane by changing the rules of geometry? The answer is: Yes, we can. A geometry of non-Euclidean kind can be found *to incorporate earth's gravity* in which the trajectory (35) is the *equivalent of a straight line*. We will elaborate on the expressions in italics next.

To incorporate gravity in spacetime, Einstein's proposal was to change the geometry to a non-Euclidean one. The parameters that describe a non-Euclidean geometry contain information about spacetime curvature. For a simple example of curvature, consider the two-dimensional surface of a

sphere. The geometry on the sphere is non-Euclidean; it can be characterized by the radius of curvature of the sphere or its reciprocal, the curvature of the sphere. If some such parameter can be interpreted as describing the gravitational effects, then we have incorporated gravity into geometry.

The straight line in Euclid's geometry is the line of shortest distance. To seek its equivalent in a non-Euclidean geometry consider again the example of the sphere. Take any two points P_1, P_2 on its surface and stretch a thread across them. If we pull the thread slightly and let go, it comes back to its original position. It can be easily verified experimentally and proved by calculus of variations that this position lies on the great circle passing through P_1 and P_2. Thus by requiring the length of the curve joining P_1 and P_2 to be an extremum we can determine a unique path joining them. This is the equivalent of a straight line.

Because any two great circles on a sphere intersect, there are no parallel lines in this geometry. Thus Euclid's parallel postulate breaks down. A triangle on the sphere, as shown in Figure 4, has its three angles adding upto greater than two right angles. These are indications of a non-Euclidean geometry.

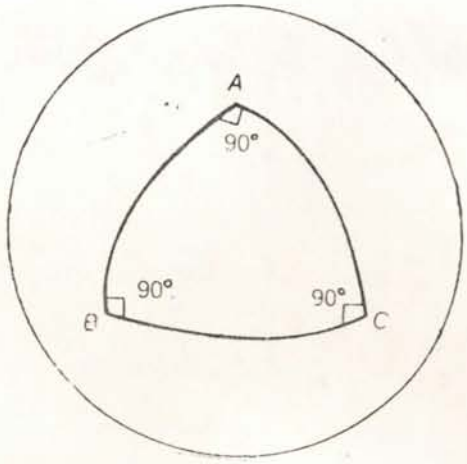

Fig. 4. The triangle ABC on the spherical earth with A the North pole and B, C on the equator on merideans 90° apart, has its three angles adding upto *three* right angles.

10. Einstein's Equations

Einstein realized that giving a quantitative expression to the above concept required knowledge of Riemannian geometry. It took him some time and help from colleague and friend Marcel Grossmann to master the subject and some trial and error before arriving at the final form of field equations for general relativity. Symbolically, the equations are:

$$G_{ik} = -\kappa T_{ik}. \tag{37}$$

where κ is a coupling constant, T_{ik} the energy-momentum tensor of all matter and radiation that is present in spacetime and G_{ik} is a tensor describing spacetime curvature. Both sides of the equations can be written out fully as ten partial differential equations in four independent variables x^i ($i = 0, 1, 2, 3$) and ten dependent variables g_{ik} which constitute the spacetime metric tensor.

The role of g_{ik} can be seen as the generalization of η_{ik} of special relativity. If x^i are the spacetime coordinates, the proper distance ds between two neighbouring points with coordinates x^i and $x^i + dx^i$ is defined as the quadratic form

$$ds^2 = g_{ik}\, dx^i\, dx^k \tag{38}$$

(Note: the summation convention is used). The g_{ik} are functions of spacetime coordinates. Given the physical information on the right hand side of (37), we can determine the spacetime geometry by solving those equations.

The second important content of the Einstein equations is the equations of motion. We have already seen that gravity has been eliminated as a force—its effects are to be felt through spacetime geometry. So as per Newton's first law of motion, a particle moving in a gravitational environment is in fact moving under no force. As such it should move in a straight line with uniform velocity. This notion in curved spacetime is described by a 'geodesic'. Mathematically, a geodesic joining two points in curved spacetime is determined as the *line of stationary length* between them.

The length of a line Γ connecting two points A, B is simply the integral

$$L = \int_{\Gamma} ds. \tag{39}$$

If we 'vary' the curve Γ slightly, i. e., take a neighbouring curve Γ' joining A and B, stationarity implies that to first order the length of Γ' should be the same as the length of Γ. Thus we have to solve the problem from the calculus of variations:

$$\delta L = 0. \tag{40}$$

This requirement gives us the equations of the geodesic joining A and B.

The statement that a geodesic describes motion under no force can be physically interpreted through a general principle that underlies general relativity. As we saw in our opening example of a projectile moving under the earth's gravitational environment, the parabolic world line can be looked upon as a geodesic. The relationship between (35) and (36) can be made more obvious if we define a new space variable

$$y' = y + \frac{1}{2} gt^2.$$

The origin of the y-axis then corresponds to a point accelerated upwards in the y'-frame. In other words, a rocket suitably accelerated upwards can simulate a gravity-like effect for anyone riding on it. This equivalence between gravity and accelerated frames is called the *principle of equivalence* and it is supposed to extend to all of physics.

Stated generally, this principle means that physics in the presence of gravity at any given world point can be simulated by that in a suitably accelerated frame. The equivalence is strictly local and applies only to a small neighbourhood of the point in question. In the example of the projectile the value of g changes with height and so the equivalent accelerated frame cannot be found at all points. In Einstein's celebrated example, the man in a freely falling lift feels weightless—because the acceleration has cancelled the earth's gravity. Here too the effect is strictly local. There is no lift that eliminates the earth's gravity everywhere and at all times.

The above principle tells us that in order that such accelerated frames exist the inertial mass of any particle must equal its gravitational mass. For, if this were not the case different particles would respond differently to a given gravitational environment and describe different world lines and we could not find a unique accelerated frame for all. This argument is a restatement of Galileo's celebrated law of falling bodies.

The general theory of relativity is exceptional in the class of physical theories because of this feature: its field equations contain the equations of motion. In other theories like the electromagnetic theory, this is not the case; Maxwell's field equations have to be supplemented by the equations of motion.

We next consider a simple solution of the complicated Einstein equations (37).

11. The Schwarzschild Solution

What is the geometry of spacetime in the vicinity of a massive spherical object? This question was answered by Karl Schwarzschild in 1916. The Schwarzschild solution has played a crucial role not only in the conceptual understanding of relativity but also in testing the theory experimentally.

Let us first consider the geometry of empty spacetime. If we choose spherical polar coordinates (r, θ, ϕ) to describe spatial geometry and t as the time coordinate, then the Minkowski rule (15) can be rewritten in a differential form as

$$ds^2 = c^2\, dt^2 - dr^2 - r^2\,(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{41}$$

Now we consider the gravitational effect of putting in a sphere of mass M with centre at $r = 0$. How is (41) changed? By solving Einstein's equations, Schwarzschild found the answer to be:

$$ds^2 = c^2\left\{1 - \frac{2GM}{c^2 r}\right\} dt^2 - \frac{dr^2}{1 - 2GM/c^2 r} - r^2\,(d\theta^2 + \sin^2\theta d\phi^2). \tag{42}$$

How did we acquire the Newtonian gravitational constant G in the above

solution? The field equations (37) do not contain G; instead they contain a coupling constant κ.

The relationship between κ and G is established only by a comparison of Newtonian theory with Einstein's in the limit of *weak fields*. Thus from (42) we see that for $r \gg 2GM/c^2$ the differences between (41) and (42) are small. The geometry of spacetime is only marginally non-Euclidean. It can be shown that in such a case Einstein's equations approximate to Newton's provided

$$\kappa = \frac{8\pi G}{c^4}. \tag{43}$$

Thus in the original solution of (37) for the Schwarzschild case one would obtain the coefficient of dt^2 to be $c^2(1 - A/r)$, with A an arbitrary constant. The identification of A with $2GM/c^2$ is made by referring to the above weak Newtonian limit.

By working out the equations of geodesics we can determine how test particles move in the gravitational influence of the central mass. An important result to emerge is that the motion of a planet around the sun can be so described. It agrees with the Keplerian motion in the weak field limit. There is one important difference, however. The perihelion of the orbit slowly precesses. The effect is the largest for the innermost planet Mercury whose perihelion would precess according to the theory, at the rate of 43 arc sec per century. In actual fact Mercury's perihelion precesses at the rate of 575 arc sec per century of which 532 arc sec per century is simply due to the Newtonian gravitational effect of other planets. Thus the discrepancy between observations and Newtonian theory is exactly accounted for by the above relativistic effect. This effect for outer planets is negligible.

General relativity, however, gained worldwide recognition because of the verification of its prediction that even light rays are bent by gravity. This comes about in the following way. The generalization of (42) from (41) together with the fact that light speed is c leads to the conclusion that light ray is a null geodesic. That is, apart from (40) the light ray also satisfies the condition

$$ds = 0. \tag{44}$$

Using the geometry of Schwarzschild's spacetime one can compute the path of a light ray grazing the limb of a massive object. The path is different from the Euclidean straight line path and this can be detected astronomically as follows.

Suppose we observe a star when its line of sight grazes the solar limb, and also at other times when the sun is nowhere near the line of sight. We are then effectively observing the null geodesic direction to the star for (42) in the first case and (41) in the second. Calculations show that the change in apparent direction of the star in the two cases is

$$\alpha = \frac{4GM}{c^2R} \cong 1.75 \text{ arc sec} \tag{45}$$

where M is the mass of the sun and R the radius of the sun.

The original observations were taken by Eddington and others on the occasion of the total solar eclipse of 1919. During an eclipse it is possible to see the star just 'crossed' by the solar limb. The predicted bending was found although the inherent error bars for the observations were much larger than estimated at the time. The observations have since been repeated using microwaves rather than visual light. The change in the apparent displacement of the direction to the quasar 3C 279 was found to be in agreement with general relativistic prediction within 1% error.

The time delay in the return of a radar signal bounced off the earth back to spacecraft Mariner 6 and 7 when it grazed the solar limb was also found to be in agreement with the predicted value of about 200 μ second. Thus non-Euclidean effects in space-time geometry are not only measurable but have actually been measured.

The simplest of these effects is the observed change (increase) in the wavelength of light travelling from the stronger gravitational field of a white dwarf star to us (in a weaker gravitational field). Known as the gravitational redshift, the order of magnitude of this effect is $\sim 10^{-4}$. In general, the quantitative effect is $\sim GM/c^2R$, where M is the mass and R the radius of the star.

The general theory of relativity has been tested and found to be satisfactory in such cases of weak gravitational fields. The principle of equivalence and its consequence of the equality of inertial and gravitational masses has also been tested to a high degree of accuracy both astronomically and in the laboratory. The theory remains to be tested in two important areas, however.

First, we have to devise suitably sensitive machinery to detect gravitational radiation. The theory predicts that such radiation comes from a source whose quadrupole moment is changing with non-zero time derivative of the third order. The effect is extremely small and beyond detection by foreseeable technology for man-made sources in the laboratory. In this respect the situation is far worse than for the detection of electromagnetic radiation. The hope lies in detection of radiation from a cosmic source like a supernova or, less dramatically, a binary star system.

An indirect way of measuring the gravitational radiation coming from a binary star system is to look for gradual decrease in its period. This is expected to happen because as a result of energy loss, the system shrinks steadily with its components going round each other more and more rapidly. Such an effect has been found for the binary system one of whose members is a pulsar PSR 1913 + 16.

The second important area in which general relativity remains untested is in the case of strong gravitational fields. In the Schwarzschild solution the spacetime geometry of (42) becomes more and more different from

Euclid's of (41) as we approach the surface of the body and as its radius itself becomes smaller and smaller. There are two important stages that arise if a body is contracting under its own gravity. The first stage is when its radius becomes equal to the so called Schwarzschild radius

$$R_s = \frac{2GM}{c^2}. \tag{46}$$

Computations of null geodesics will show that at this stage no light signals will reach an outside observer at $r > R_s$. At such a stage the body is given the name of a *black hole*.

The second stage is reached when the entire body shrinks to a point $r = 0$. This is the so called *singular stage* when the mathematical description of spacetime geometry entirely breaks down. Are both these stages too esoteric to be of physical relevance? For this is where the full effect of a non-Euclidean geometry would be observable.

Theoretically speaking, such stages could be reached if a massive body continues to collapse under its own force of gravity. A stage like this may arise for a very massive star at the end of its evolution when all its nuclear energy sources are exhausted. A star like this with mass exceeding three times the mass of the sun cannot support itself in equilibrium against its own gravity by any known physical effect. Do black holes get formed as a result of this? Do they continue to collapse to a singular state, as general relativity requires them to? These questions have not yet been tested experimentally or observationally.

12. Cosmology

We briefly refer to another fascinating application of general relativity. Unlike the Schwarzschild solution which is concerned with an isolated gravitating mass we now consider the entire universe. The large scale structure of the universe is found observationally to be fairly homogeneous and isotropic. Can we solve Einstein's equations to determine the geometry of spacetime *as a whole*? The problem was first tackled by Einstein himself in 1917, who looked for a model universe that was by and large static. However, a series of dynamic solutions obtained by Alexander Friedmann in 1922 turned out to be more realistic.

For, in 1929 observations by Edwin Hubble and Milton Humason confirmed that there is a systematic increase of wavelength in the light received from galaxies and clusters of galaxies. This effect, known as *cosmological redshift* z was found by Hubble to increase with the distance D of the source:

$$z = HD/c. \tag{47}$$

The constant H is today called Hubble's constant. Friedmann's dynamic

models could account for the effect naturally, whereas Einstein's static model failed in this respect.

The Friedmann models describe a universe that is *expanding*, i.e., the volume containing any given set of galaxies is steadily increasing with time. The models also show that a finite time ago this volume was zero. That instant, again one of singular geometry as in the black hole case, is identified with the beginning of the universe. The tremendous burst of activity soon after the beginning gives the epoch the popular name of *big bang*.

The homogeneity and isotropy of the universe determine a unique frame at each world point. An observer moving relative to that frame will not see the universe to be isotropic. It is interesting to note that this unique frame is *inertial*. This was noted by Ernst Mach back in the last century. Mach interpreted this result to imply an intimate connection between the property of inertia and the large scale structure of the universe. This concept is known as Mach's principle. It also provides the necessary input to the twin paradox: the distinction between inertial and non-inertial twins can now be made by viewing the universe!

It was Einstein's hope to demonstrate that Mach's principle is part of general relativity theory. This hope was not realized, however, and it is now understood that to incorporate Mach's principle, one needs a wider framework than general relativity.

General relativity may nevertheless rightly claim the credit for launching cosmology on the scientific scene. The geometry of the large scale structure of the universe may well be non-Euclidean; but various attempts to test this hypothesis have so far been inconclusive. This is largely because the measurements needed for the purpose take us to distances of billions of light years and to time scales of several billion years; and these measurements involve observational errors that exceed the range of geometrical variations to be tested. On the other hand, the physical aspects of cosmology have received increasing attention in recent times; but their discussion would carry us well beyond the scope of this article.

III. Conclusion

To summarize, the special theory of relativity has now become an essential foundation of modern physics. The measurements of physical quantities in any theory or experiment have to be consistent with this theory. The general theory, though aesthetically and intellectually on a higher plane than the special theory, has limited applications and it has been tested fully only in its weak field limit. By doing away with gravity as a force, the theory has somewhat isolated itself from the mainstream of physics although it continues to be the best working theory of gravity. The

problem of interpreting gravitational radiation has been a hard one and that of quantizing gravity is still with us in this theory. It may well be that a better understanding of gravity will come only through its unification with the rest of physics.

References for further reading

Special Relativity for Physicists by G. Stephenson and C.W. Kilmister (Longmans. Green and Co, 1958)

General Relativity and Cosmology by J.V. Narlikar (Macmillan, 1978).

2

QUANTUM MECHANICS

P.M. AGARWAL
School of Studies in Physics
Vikram University, UJJAIN 456 010

1. Black-body radiation
2. Photoelectric effect
3. Bohr's theory
4. Dual nature
5. Heisenberg's discovery
6. Schrodinger's creation
7. Hydrogen-like atom
8. Intensity and selection rules
9. Zeeman effect, Paschen-Back effect and Stark effect
10. Atoms having more than one electrons
11. Molecules
12. Operators in quantum mechanics
13. Eigenvalues, eigenfunctions and uncertainty principle
14. Questions and Problems.

Quantum mechanics is one of the best creations of human mind. As its history is very helpful in developing clear concepts, we would here introduce quantum mechanics after describing the historical development.

1. Black-Body Radiation

We all are aware of light. Light is a form of electromagnetic radiation. Different colours of light are due to different values of wavelengths of light. The wavelength of visible light ranges from 4000 Å to 8000 Å. Thus light is nothing but electromagnetic radiation in the range 4000 A to 8000 Å. There exist electromagnetic radiations beyond this range also. Electromagnetic radiations of wavelengths greater than 8000 Å are infrared, microwaves, and radio waves. Ultraviolet rays, X-rays and gamma rays are electromagnetic waves having wavelengths smaller than 4000 Å. If we arrange all these forms of waves in increasing order of wavelength then