World-Structure and the Expansion of the Universe.

By **E. A. Milne.**

With 6 figures. (Received November 14, 1932.)

The phenomenon of the expansion of the universe is shown to have nothing to do with gravitation, and to be explicable, qualitatively and quantitatively, in terms of flat, infinite static Euclidean space. A new method of investigation of the cosmological problem is developed on the basis of two and only two postulates, without any appeal to causal concepts, laws of dynamics, curved metrics or gravitational theory whether of the type of field theory or action at a distance. The first postulate is Einstein's postulate of the constancy of the velocity of light; it sums of those experiments which failed to detect "motion through the aether" and it is equivalent to the statement that two observers in uniform relative motion have *consistent* views of events when their separate descriptions are connected by the Lorentz transformation. The second postulate, though connected with a principle stated by Einstein, is fundamentally new and asserts that two observers in uniform relative motion have *identical* views of the universe, i. e. that each sees the same evolving sequence of world-pictures, in his own Euclidean space and his own time-scale. These two postulates are shown to imply Hubble's velocity-distance proportionality for the spiral nebulae. Further they lead to a determination of the distribution of matter and motion in the universe. According to these distribution-laws the universe of nebulae occupies the interior of an expanding sphere (in 3-dimensional Euclidean space) whose centre is in any arbitrary nebula of the system. The outer boundary of the sphere expands with the velocity of light, and the nebulae, infinite in number, are strongly concentrated towards the expanding boundary; but the total light received in any direction is finite. These statements hold in the Euclidean space and ordinary time of the observer moving with the arbitrarily chosen nebula which is the centre of the system for this observer. The above description of the universe may be regarded as an extension to material systems of Einstein's analysis of expanding light-spheres. Every nebula of the system is subject to zero acceleration, and this fact amounts to a new statement of the law of gravitation, expressed in terms neither of a metric nor of action at a distance, but as a total effect of the material of the universe at determinate points.

I. *Kinematic solution and comparison with current Theories.*

§ 1. Scope of the paper. 1. The object of this paper is primarily to present a simple kinematical explanation of the phenomenon of the expanding universe. In contrast to the current theory which involves the notion of curved, finite, "expanding" space, it is shown that the facts

can be accounted for kinematically by adopting flat, infinite, static Euclidean space as the seat of natural events. Further, whilst the current theory attributes the expansion phenomenon to an effect of gravitation (field-equation theory) the new theory shows that the expansion is in reality a natural phenomenon taking place in the absence of gravitation and proceeding in spite of gravitation. Expansion is the inevitable characteristic of the motion of an un-enclosed system possessing at some epoch concentration towards one particular region, in the frames of space and time of any one observer.

2. It is therefore held to be the logical procedure to examine the nature of the expansion, in the first instance, in the light of the so-called "restricted" theory of relativity and the examine later the modifications introduced by gravitation. By this means I show that the apparently contra-relativistic idea of a concentration of matter towards one particular region of space is an effect created by the observer. An observer is necessary in order to describe "what is". The presence of the observer necessarily differentiates out particular properties of the spatio-temporal distribution of matter and motion, which properties bear special relationships to the observer. The world then appears to be centred round the observer wherever he be, provided he assumes a velocity equal to that of the matter in his own neighbourhood, i. e. provided he chooses his frame of reference so that he is at rest with respect to his immediate surroundings. The world is then perfectly ego-centric at all points and its unfolding history appears the same from all points. This solution satisfies the condition which EINSTEIN postulated that any solution of the cosmological problem must satisfy: „*Alle Stellen des Universums sind gleichwertig*"[1]) but it is essentially different from the uniform-density, zero-velocity type assumed by EINSTEIN when he went on to say „im speziellen soll also die örtliche gemittelte Dichte der Sternmaterie überall gleich sein"[1]). The distribution of matter and motion in the ground-plan of the universe here obtained is shown to avoid those difficulties which originally led EINSTEIN to adopt a curved finite continuum for the world. It is somewhat remarkable that it had been previously assumed without proof that the only possible density-distribution satisfying EINSTEIN' postulate was a uniform one filling a finite continuum. Actually we can construct non-uniform distributions of matter and motion which bear the same relation to the observer wherever he be. These distributions exhibit the principal large-scale phenomena of the universe: expansion with a velocity-distance proportionality, and a velocity-distribution

[1]) Sitzungsber. d. Preuß. Akademie d. Wissenschaften 1931. S. 235.

which when modified by the effects of gravitation appears to give rise to the observed agglomeration of matter into closed self-contained galaxies.

3. The method here employed for developing the kinematic explanation is essentially an "astronomical" one. That is to say, it endeavours to discover the laws of nature as far as they concern gravitation and dynamics by studying the distribution of matter and motion capable of existing in the world, in the large, when a modified form of EINSTEIN's postulate is imposed; it proceeds by observation of what is there to observe. This may be contrasted with the physicist's procedure which is to discover the laws of nature by experimentation, micro-examination of specifically devised situations. The method of the present paper[1]) is the examination of the macroscopic state of the most general possible world consistent with the ideas of relativity. Since the only things we actually observe in the world are positions and velocities at known instants in suitably selected frames, the subject of our study is simply matter and motion — the counting of particles and the ascertaining of velocities in the space-frame and according to the time-scale of any one observer. The observer is supposed to be unaware of any dynamical laws and to have no opportunity of making experiments, but to be attempting to frame the laws of nature governing what is observed; he is to be KEPLER and NEWTON combined without GALILEO. Notions such as mass, momentum, energy, force do not enter. The method is kinematical and not dynamical and dynamical concepts play no part, for dynamical concepts are mere convenient modes of thought — labels introduced by the analyst — and are in the first instance unnecessary in an "astronomical" description. Dynamical concepts only become necessary when we wish to carry over laws of nature, astronomically ascertained, to other physical phenomena in arranged experiments or to other astronomical situations. When the astronomical situation studied is the world itself there are no other astronomical situations.

This general object is only partially achieved in the present paper. The following results are however obtained.

4. An essential step is the introduction[2]) of the observer as a necessary element in the situation. EINSTEIN's postulate that all places in the universe must be equivalent is modified to read: *The universe must appear the same to all observers.* As the view of any particular observer depends on the space-time frame be adopts, this requires to be made more precise. We therefore posit: *not only the laws of nature, but also the events occurring*

[1]) See Part II.

[2]) Following EINSTEIN.

in nature, the world itself, must appear the same to all observers, wherever they be, provided their space-frames and time-scales are similarly oriented with respect to the events which are the subject of observation. By "the world" I do not mean "the world at any instant" but the totality of the flux of events. This postulate is referred to briefly as the "extended principle of relativity".

It is shown to be a consequence of the postulates: (1) that a world without gravitation (i. e. a world in which all the paths are straight lines described with uniform velocity) cannot exist without singularities[1]), i. e. that gravitation is a necessary constituent of the real world; (2) that in the real world there is a connection between light and gravitation, in the sense that the propogation of light along rectilinear paths with constant (observable) velocity through the real world is incompatible with the existence of a real world free from singularities.

There is of course nothing new in these results as facts. The novelty lies in their derivation, by statistical analysis, from the postulates of special relativity. The mathematical methods employed are also novel. No assumptions are made of a dynamical character, such as that the path of a material particle is a geodesic in space-time or that the path of a light-ray is a nul-geodesic.

5. It is held throughout to be a tenable position that space and time or space-time on the one hand and laws of nature of the other hand are no objective entities with an objective existence; and that it is impossible to speak objectively of the "curvature of space" or the "law of gravitation". EINSTEIN and MINKOWSKI showed that space and time are merely different resolutions, by different observers, of that reality which is the change or flux in the spatio-temporal relationships of events. Similarly in this paper it is considered that space and time (together) on the one hand, and laws of nature on the other hand are merely different resolutions of reality by different thinkers — complementary aspects of reality[2]). Either Newtonian space and time may be posited for the universe, and the law of gravitation then determined; or alternatively the principle of equivalence and a law of gravitation may be posited, and the curvature of space-time then determined.

6. The above train of ideas arose out of a study of the problem of the observed expansion of the spiral nebulae. In pursuit of this original problem I

[1]) These singularities are all at a finite distance from the observer. There is no singularity at infinity.

[2]) I do not imply that in all cases the resolution can be made orbitrarily; I mean only that a certain freedom of choice exists.

determine in detail the statistical distribution of matter and motion in the universe that would be implied by the postulate of the constancy of the velocity of light and the postulate of the extended principle of relativity. Although no general theory of gravitation is attempted[1]), the probable effect of gravitation in modifying this distribution is outlined. It is shown that the singularities would probably become the seat of discrete agglomerations of matter receding from a common centre (in the frame of any arbitrary observer), each separate system being "gravitationally closed" but the systems being (roughly speaking) gravitationally independent of one another and escaping from one another. A rigorous solution embodying these ideas is given in § 13. The solution and indeed the whole theory involve consideration of the problem of whether there was in the astronomical past a "beginning of things".

7. To forestall misunderstandings it may be worth while to state explicitly that the paper is in no sense a criticism of EINSTEIN's general theory of relativity although it avoids the use of it. EINSTEIN's special theory of relativity in its kinematical aspects is accepted *in toto,* and used freely. The paper does however criticize EINSTEIN's and de SITTER's statical solutions of the cosmological problem. And it criticises also, on physical grounds, all current "non-static" solutions of the cosmological problem. These criticisms are not introduced in any hostile spirit, but simply to make clear the relation of the new treatment to the existing treatments.

8. A summary of an early version of the paper has been published in "Nature" 1932, July 2, to which readers are referred. See also "Nature" 1932, Oct. 1.

Although the work was substantially carried out at Oxford in May 1932, I wish to express my appreciation (a) of leave of absence granted by the University of Oxford (b) of assistance from the Rockefeller Trust (c) of the hospitality of the Einstein-Institut at Potsdam, all of which have contributed to the leisure necessary for the completion of the paper. I am also personally indebted to Prof. E. FINLAY FREUNDLICH for his numerous sympathetic and helpful criticisms and for his putting at my disposal his intimate knowledge of relativity literature[2]).

---

[1]) But in §§ 13, 14, written after this introduction had been completed, an exact solution is arrived at which is equivalent to a new formulation of the law of gravitation.

[2]) A discussion of the kinematic theory written by Prof. E. F. FREUNDLICH, from a slightly different point of view, will appear in Naturwissenschaften, 1933, Heft 4.

§ 2. The observational facts. 9. It was discovered by V. M. Slipher in 1912 and confirmed in 1922[1]) that the spiral nebulae possess radial velocities (as determined spectroscopically) which are predominantly velocities of recession. It was found further that their velocities are large compared with the ordinary range of stellar velocities. Slipher's researches have been confirmed and greatly extended by Hubble and his collaborators. The Mount Wilson observers have determined spectroscopically the radial velocities of a great number of spiral and other galactic nebulae and in almost all cases the velocity is away from ourselves. Moreover the velocity of recession was found to be greater the greater the distance, and in 1929 Hubble[2]) enunciated the law that the mean velocity of recession at a given distance is proportional to the distance. Graphs illustrating this law have been given by Hubble (loc. cit) and by de Sitter[3]). It is not yet known with what accuracy the individual recession-velocities obey this law, but it seems probable that the "proper motions" of the nebulae (i. e. the deviations in radial velocity from the velocity-distance proportionality) are relatively small, according to J. H. Oort[4]) of the order of 80 kms. sec.$^{-1}$. This number is not however well-determined owing to uncertainties in the assigned distances. More recent results are given in a paper by Humason[5]), where the measured velocities range up to 20000 kms. sec.$^{-1}$, or $^1/_{15}$ of the velocity of light[6]). The usually adopted constants for Hubble's law are equivalent to a recession-velocity of approximately 500 kms. sec.$^{-1}$ per 10^6 parsecs.

10. Enormous numbers of faint extra-galactic nebulae have been photographed with the world's largest telescopes. They have been partly counted, and their magnitudes have been estimated, but they have for the most part not yet been observed for radial velocity. It seems highly probable that the fainter the apparent magnitudes the more distant the nebulae[7]), and hence if Hubble's law or any thing like it continues to hold

---

[1]) V. M. Slipher, *Lowell Observatory Bulletins.*

[2]) E. Hubble, Proc. Nat. Acad. Sci. **15**, 168, 1919.

[3]) B. A. N. No. 185, 1930, p. 167.

[4]) B. A. N. No. 196, 1930.

[5]) Mt. Wilson Contr. No. 426, 1931.

[6]) Prof. de Sitter has kindly informed me that still higher velocities have since been observed.

[7]) But a nebula receding with a velocity comparable with that of light suffers an appreciable diminution of brightness to an observer at rest, and will in fact be much nearer than would be estimated from its apparent brightness uncorrected. See concluding sections of Part II for quantitative treatment; also appendix.

for a further factor of even 10 in distance, nebulae must exist with recession-velocities comparable with that of light. If this anticipation is confirmed by observation, the fact is of fundamental theoretical importance, for it means that no frame of reference exists in which the velocities of celestial objects are all small compared with that of light. This removes one of the arguments which led to EINSTEIN's spherical world.

The well-known phenomenon we have briefly described is usually referred to as the "expansion of the universe"[1]).

§ 3. Kinematic explanation of the expansion phenomenon. 11. We first examine an abstract problem. Consider the kinematics of a swarm of particles which move in straight lines, each with a uniform velocity, without collisions or other interactions. Let them possess an entirely arbitrary (not necessarily random) velocity-distribution the velocity-distribution not being necessarily the same at different points. In the Euclidean space and Newtonian time of any given observer, let the particles have a density distribution which is such that at a given time $t = 0$ the particles occupy a sphere S of radius r_0, centre $x = 0$, $y = 0$, $z = 0$. Outside the sphere space is to be empty; inside the sphere the particles may have any density distribution whatever.

Let as examine the subsequent motion of the system. At any point of the surface of S, the *outward moving* particles will move into the empty space outside. The faster moving particles will gain on the slower, and at any later time t the fastest moving particles will form an expanding spherical frontier-zone, which will be followed by and partly overlapped by the next fastest, and so on. The *inward moving* particles will move inwards, traverse a chord of the sphere, emerge at the other side and then move outwards. The same will hold good for any starting point. After the lapse of sufficient time *all* the faster moving particles will be moving outwards, and their velocities will be predominantly motions of recession, approximately directed away from the centre $x = 0$, $y = 0$, $z = 0$. Only the very slowly moving ones will then be moving inwards, namely those which started in an inward direction, not too near the centre and are still approaching the centre.

[1]) The use of the word "universe" has come to be ambiguous in this connection. The word "universe" may refer to the system of the spiral nebulae, or to the totality of material things, or to some hypothetical "existing-in-itself" framework called space. As it is the standpoint of the present paper that "space" by itself has no existence, I shall use the phrase "expansion of the universe" to mean simply the observed expansion of the system of the spiral nebulae.

Provided the original velocity-distribution was a *continuous* one and included the velocity *zero* the original sphere will always be occupied, however long the time; for there will always be some slow-coaches which have not had time to traverse the sphere S.

At greater distances from the centre than r_0 a sorting process goes on, the particles dividing themselves into spherical zones of gradually increasing recession-velocity as we go outwards. The mean velocity of recession at distance r at time t is approximately r/t. For for t sufficiently large, the distance from the origin of a particle of velocity V is approximately V. To see this in more detail, consider the sub-group of particles which are moving with given vector-velocity V. At time $t = 0$ they occupy a sphere centre O, of radius r_0. At time $t > 0$ the sphere of particles will have moved a vector-distance $R = Vt$. For $t > r_0/|V|$ this sphere will be between the spheres centre O, of radii $|V|t - r_0$ and $|V|t + r_0$. Thus the actual distance r of any particle satisfies

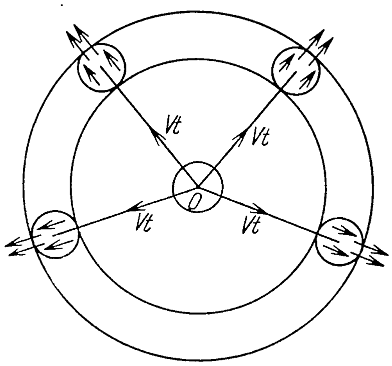

Fig. 1.

$$|V|t - r_0 < r < |V|t + r_0$$

which gives

$$\frac{r - r_0}{t} < |V| < \frac{r + r_0}{t}$$

or ultimately

$$|V| \sim r/t.$$

Thus statistically we have an expansion phenomenon, a systematic recession for all save very near particles and a velocity-distance proportionality law. The velocities themselves remain constant, and the coefficient of proportionality is simply the reciprocal of the epoch: consequently the coefficient of proportionality decreases as the time increases.

It is obvious that similar results will hold for less specialised initial density-distributions. There is no necessity for the system to be confined initially to the interior of a sphere. It is merely necessary that the initial density distribution shall decrease with the distance at sufficiently large distances, i. e. that the initial density-distribution shall be originally concentrated towards a particular region of finite extent[1]). We shall then

[1]) More rigorously, it is necessary that the sub-system for each velocity V shall have this property.

have an expansion phenomenon, a recession predominance and a velocity-distance proportionality. For the process is essentially one of diffusion. Any irregularity in density will tend to smooth itself out by expansion. The result holds good even if the system originally fills infinite Euclidean space. Sorting-out of velocity goes on, and though there may be a few fast particles arriving in the vicinity of the origin from distant regions of space, these will not be numerous and the faster moving particles will tend to be found at the greatest distances from the origin. Collisions will retard the process of the diffusion, but not seriously affect it, and the same is true of any other kind of interaction which is not so powerful as to keep the particles of the system for ever in proximity to one another, i. e. which is not so powerful as to prevent "escape to infinity". The essential aspect of the problem is that the system is *unenclosed* and has an infinite continuum in which to expand.

12. This kinematic behaviour of a cloud of particles strongly resembles the observed behaviour of the extra-galactic nebulae. It is so simple and direct as an explanation of the expansion phenomenon that it appears to be worth while to examine it in further detail. We have simply to suppose that spirals in general are endowed with velocities exceeding the velocity of escape from the rest of the system. They will then pursue paths practically indistinguishable from their rectilinear asymptotes. Collisions between spirals we may disregard in a first survey. The system of spirals is then completely analogous to the system of particles just considered, and the same results follow, namely the recession phenomenon and the velocity-distance proportionality, provided only, in the frame of any one observer, there is an excess concentration at some epoch towards some particular region. The coefficient k in the law $v = kr$ is simply $1/t$, where t is the epoch of the present moment measured from the epoch of a certain configuration in the past. Hubble's law gives in this way

$$t = 0.6 \times 10^{17} \text{ secs} = 2 \times 10^{9} \text{ years}.$$

The coefficient k is not a constant; on the contrary it may be expected to decrease secularly as time goes on; the distances r increase but the velocities remain approximately constant[1]).

---

[1]) The recession law $\frac{dr}{dt} = \frac{r}{t}$ markedly differentiates the kinematic explanation from the "expanding space" explanation, where the recession law is $\frac{dr}{dt} = kr$, k being constant in time as well as in space.

13. This explanation is so elementary that it has hitherto escaped notice. Had it been suggested before general relativity was known doubtless it would have been accepted as obvious. As it is it will have to meet the accumulated momentum of fifteen years' work on apparently more recondite lines.

A fundamental difference between explanations of the current "expanding space" type and the kinematical explanation, is that the former attribute the observed expansion to gravitational influences, whilst the latter views the expansion as occurring *in spite of* gravitational influences. The preference for the kinematical explanation is immediate on grounds of simplicity. For expansion (given an un-enclosed system with an initial region of concentration) occurs inevitably whether gravitational forces are present or not. It is therefore in accordance with the principle of scientific economy to retain the kinematical explanation so long as may be possible.

But a second reason for preferring the kinematical explanation is that it gives at once a positive expansion, whilst on gravitational theories it has hitherto been impossible[1]) to decide whether expansion or contraction is to be expected. The inevitability of the positive sign for the expansion on the kinematic explanation is a strong reason for preferring it.

Since the Newtonian law of gravitation holds to a very close approximation in the space and time of any one observer, it should be possible to examine the effect of gravitation in modifying the details of the expansion on the Newtonian theory alone. A relativistic theory of gravitation should not be necessary. We can therefore on a first survey examine the kinematic explanation using the Newtonian law of gravitation, Newtonian time and Euclidean space. When we pass to a second observer, to a good first approximation we should expect that the special theory of relativity and the Lorentz transformation should be sufficient[2]).

14. The kinematical explanation at once raises certain fascinating questions. They are of two types, one type being those connected with the space and time of one observer, the other type being connected with the relationship between the observations of two different observers. Questions of the former type are.

(1) why is every spiral already endowed with a velocity exceeding the velocity of escape?

---

[1]) Cf. Council Note, M. N. R. A. S. Feb., 1931 (A. S. E.).

[2]) Actually, the final solution found in §§ 13, 14 is rigorously exact on the Lorentz transformation, whatever the law of gravitation.

(2) what is the significance of the apparent existence of an initial configuration and an origin of time?

Questions of the latter type are:

(3) how can the apparent existence of a preferential region of space, namely the region of the initial concentration, be reconciled with our general ideas of the relativity of space?

(4) what is the relation between the time t which occurs in the coefficient in the velocity-distance law, measured in the time-scale of one observer, and the other time-scales that would be used by other observers in motion with respect to the first observer?

The rest of the paper is chiefly concerned with seeking answers to these questions.

The answers are not independent. It will be shown by means of the special theory of relativity, with the additional postulate of the extended principle of relativity, that we are to expect for the ground-plan of the universe a distribution of matter and motion which, in an appropriately chosen space-time frame depending on the observer, has the same appearance to every observer where ever he be in the frame of any other observer. The answers to (3) and (4) are thus thus that the apparently preferential region of concentration has no "absolute location" in space for there is no absolute space in which to have a location, but is created by each observer by his own deliberate choice of a frame of space-time.

It will be further shown that the distribution of motion in the ground-plan of the world corresponds to the existence of velocities of all values up to c the velocity of light. It is only this type of distribution which is consistent with the non-existence of a preferential velocity-frame. This suggests immediatly the answer to (1). Those particles which cannot escape from a given particle P_1, form a closed system S_1 round P_1. Those which have a velocity exceeding the velocity of escape from P_1 simply escape. Let P_2 be such an escaping particle. Of the other particles escaping from P_1 those which cannot also escape from P_2 from a closed system S_2 round P_2. Those which can escape, do. Let P_3 be a particle escaping from both P_1 and P_2. Repeat the argument for P_3 and so on indefinitely. By the special theory of relativity there will exist an infinite series of "relative escape velocities" $v_1, v_2, v_3 \ldots$, tending to c, with the property that v_n is sufficient to enable P_n to escape from all of $P_1, P_2, \ldots P_{n-1}$. The result is a system of closed sub-systems, $S_1, S_2, \ldots S_n, \ldots$ infinite in number, all escaping from one another[1]). Whether these systems are to be identified with spirals or with galaxies of spirals remains for the future. But the argument is sufficient to show that the world should consist of closed sub-systems all escaping from one another with velocities ranging up to c.

As regards the answer to (2), it will be shown that the only possible world that can exist must correspond to a system of world-lines in the time-space map possessing central symmetry about a time-space origin, $t = 0$, $x = 0$, $y = 0$, $z = 0$. This picks out a particular origin of space and a zero of time in the space and time of any arbitrarily chosen observer, but it picks out no absolute point in either space or time for neither absolute space and absolute time have any meaning. But whether the origin of time $t = 0$ is to be identified with a "beginning of things" is uncertain. All we are able to establish with certainty is that if the universe ever began, in any meaning it is possible to assign to the phrase, then it is highly improbable that (in the time of any arbitrary observer) it could begin at any other moment than the time $t = 0$ in the scale of that observer.

§ 4. Evolution. 15. Before going on to the serious mathematics of the paper, it is desirable to examine somewhat further the kinematical properties of an expanding system of particles moving in straight lines without interaction. Consider as before a single observer, with his own Euclidean space and Newtonian time, making observations on the system. By watching the particles of the system, i. e. by essentially "astronomical" observations, he can determine his own motion with respect to the system as a whole. Let him then reduce himself to rest. By determining the predominant directions of motion of the particles be can determine the locality of the original condensation, the centre $x = 0$, $y = 0$, $z = 0$. By observing the distances and velocities of the more distant particles be can compute the mean value of r/v for distant particles and so determine the value of t; and so fix (by his own clock) the epoch of the initial instant $t = 0$. The origin of space and origin of time are thus determinable from observations.

16. Now suppose that at some instant $t > 0$ the velocity of every particle is reversed. Since the system was expanding before, it must now contract. But the system must ultimately expand again, by the same arguments as before. For every particle is as before describing (in the

[1]) The argument is essentially a "DEDEKIND section" argument. We make a "section" of the system, separating those particles whose separation from a given particle P_1 has a finite upper bound from those whose separation from P_1 has the upper bound ∞.

reverse direction) a rectilinear path which stretches to infinity. Clearly the configuration at which the contraction of the system as a whole changes to an expansion is that corresponding to the instant $t = 0$. At the instant $t = 0$ itself both the system itself and the velocity-reverse of the system inevitably expand.

Thus *every* kinematic system possesses a well defined epoch $t = 0$ whether it was "initially" started at $t = 0$ or not. $t = 0$ is an epoch peculiarly associated with the system, and it will be natural to reckon all times from this epoch; it is a *natural origin of time*. It has the physical property that from $t = 0$ onwards the system naturally expands, and at $t = 0$ if the velocities of all the particles are reversed the system also expands. At epoch $t = 0$, time is unidirectional, in the sense that the system behaves in the same way whether time actually runs forwards or backwards. "Time's arrow", to use EDDINGTON's phrase, has at time $t = 0$ a barb at each end. This property holds for no other instant. For at any other instant reversal of velocities produces expansion in a contracting system and contraction in an expanding system. The epoch $t = 0$ is thus theoretically recognizable by inspection. We have simply to reverse the velocity and compare the pre-reversal motion with the post-reversal motion. If the two are indistinguishable, then the epoch of reversal must be the natural origin of time; if they are distinguishable then the epoch can at once be recognised as being either before or after the natural origin of time. These properties markedly differentiate an unenclosed system from the more familiar closed systems considered in thermodynamics[1]).

17. It is now clear if we take a distribution at *random* at a given epoch (in the language of the theory of sets of points) then save for a set of measure zero this epoch will be the natural origin of time $t = 0$ for this system. For in general a system begins to expand; the velocity distribution must thus satisfy special conditions if contraction is to occur and the reversed distribution must satisfy special conditions if contraction is to occur. Thus except in a set of measure zero, the instant $t = 0$ coincides with the instant for which the distribution is first given. Exceptional distributions are however readily constructed. For we have merely to take an arbitrary

[1]) It will be understood that I am attempting to elucidate the general characteristics of the motion of kinematical systems from physical considerations. I shall not be surprised if there exist special or exceptional velocity- and space-distributions for which it is impossible to say whether expansion or contraction is occurring; though for most systems some simple definition of expansion or contraction would probably be adequate. I commend the problem to pure mathematicians.

distribution at $t = 0$ and let it expand; then any later configuration is an example of an exceptional distribution; if reversed it contracts.

It follows that if a system is *created*, it is enormously probable (though not certain) that the instant of creation coincides with the natural origin of time for the system. If the initial density-velocity distribution were chosen at random, in all save a negligible minority of cases both the system itself and its velocity reverse would expand. But even if the system was an exceptional one it would either (a) expand immediately or else (b) contract for a limited time and then expand indefinitely. Oscillating motion never occurs; ultimate expansion is inevitable.

Thus if we take any system, we can always determine its natural time-zero $t = 0$. We need not assume that it was created at this instant; all we know is that if it *was* created, it is enormously probable that this occurred at the natural origin $t = 0$; for otherwise very special conditions would have to be fulfilled at the moment of creation. We can however trace back the motion of any system through its natural origin $t = 0$ to negative values of t and there is nothing to prevent the system *having existed* at such negative values.

18. Since expansion is inevitable, expansion is an essential constituent of evolution; it is a characteristic of the flux of reality. We may therefore say that the study of the kinematics of a simple system correlates in a single mathematical scheme evolution, creation and uni-directional time. Creation, if it takes place at all, takes place almost always (in the strict mathematical sense), at such a moment that the flight of time necessarily produces an essential characteristic of evolution. From this instant it is immaterial which way time is measured.

19. The particular application of these results to the system of the spiral nebulae is best left to the philosophical tastes of the reader. The main point is that if we consider that the spirals display the relics of the ground-plan of matter and motion in the universe as modified by gravitation, then though there is no need to postulate an instant of creation, yet the system behaves as if it had been created and moreover created at the natural origin of time $t = 0$. But nothing prevents our supposing that it was created at any negative time, or even at $t = -\infty$; it is merely enormously improbable that it was created (if it was created at all) at any epoch other than its characteristic epoch $t = 0$.

This question of creation has however only a philosophical importance. The important point is that in describing the kinematics of the present state of the universe it is natural to reckon the time-relations from the natural origin $t = 0$ and the space relations from the centre of accumulation $x = 0$, $y = 0$, $z = 0$, in the frame of the observer concerned. What "really happened" at $t = 0$ is not a scientific question.

20. If the present paper were being written before 1917 I should now go on to discuss the kinematics from the point of view of different observers with the help of the special theory of relativity. The question arises: can we determine any special characteristics of the spatio-velocity distribution likely to hold good at $t = 0$? and the *special* theory of relativity is eminently suited to this end. It may seem however to be a backward step to ignore the *general* theory of relativity and its applications to the cosmological problem. In order to justify this act of atavism I now examine the current "expanding space" theory and defer the relativistic treatment of the kinematical explanation to Part II.

§ 5. The current theory of the expansion of the universe. 21. Many attempts have been made during the last fifteen years to explain this phenomenon. It would perhaps be more correct to say that an explanation was given before the phenomenon was fully discovered, for it was predicted by a theory of de Sitter's in 1917. I do not propose to examine here the suggestions of Zwicky and Mac Millan. The explanation currently viewed with greatest favour is that "space itself", whatever that may be, is finite (but unbounded) curved and expanding. Robbed of its looseness of expression and freed from any metaphysical ideas as to what may be meant by space itself, this explanation means simply that it is held to be possible to describe the motions of the extra-galactic nebulae as a whole by assigning to them fixed co-ordinates in a continuum governed by a metric in which the spatial interval $d\sigma^2$ between two neighbouring points is a function of the time co-ordinate t, the complete continuum itself being finite, unbounded and therefore curved. The spatial interval, measured along a space-geodesic, between two non-neighbouring points of fixed spatial co-ordinates, is then a function of this time co-ordinate t, and in particular the spatial interval between the observer, supposed at rest (i. e. with fixed co-ordinates) and any assigned point (also of fixed co-ordinates) changes with the time. By a suitable choice of the metric this spatial interval can be made (under certain circumstances) an increasing function of the time, its rate of increase increasing with the interval itself. If we neglect the proper motions of the nebulae and assign fixed co-ordinates to them, we have apparently a description of the expansion phenomenon.

22. A little thought shows that this so-called explanation involves the most formidable difficulties. In the first place it has never been shown on

this theory why the universe should be expanding and not contracting. The most recent analyses[1]) envisage the possibility of *oscillating* systems in which case the observed present expansion must be attributed to an effect of phase. Earlier versions of the theory were apparently unable to determine the algebraic sign of the dilatation effect to be expected.

23. In the second place it is difficult to know what is meant physically by the co-ordinate t, whose change determines the change in the spatial interval. According to the special theory of relativity different observers moving in the world with different velocities keep different times. Which of these observers keeps the time t occurring in the metric

$$ds^2 = dt^2 - [R(t)]^2 d\sigma^2 ? \tag{1}$$

To pick out one such time is effectively equivalent to re-introducing the idea of absolute rest and absolute velocity, which is again seen in the circumstance that the nebulae are all assigned zero velocities with respect to the co-ordinate-system when they are assigned fixed co-ordinates. It was the supreme triumph of EINSTEIN's special theory of relativity that it abolished the notions of absolute rest, absolute velocity and absolute time, that in MINKOWSKI's famous (though somewhat extreme) dictum "space and time henceforth vanish and only a shadowy union of the two preserves any existence". I shall urge later that neither time nor space, either together or separately, have any "real" existence, but for the moment I am content to emphasize what has been often previously pointed out, that the assumption of a metric of the form (1) is equivalent to the introduction of "cosmic time"[2]), to the assumption of a co-ordinate t which is naturally picked out by the system of nebulae. This is equivalent to abandoning one of the physical results of the special theory of relativity.

"Cosmic time", the time occuring in the metrics of FRIEDMANN, LEMAÎTRE and TOLMAN, is the time kept by an observer "at rest" with respect to the system of nebulae, and it is with respect to this time that the continuum or space is said to be expanding. This assumes that the concept "at rest" with respect to the system of nebulae can be defined — I shall later show that it has no objective meaning. Transformations of this t into a t functionally connected with it (of the type, for example, mentioned

[1]) O. HECKMANN, Veröffentl. d. Universitäts-Sternwarte zu Göttingen, Heft 23, 1932; W. DE SITTER, Proc. Amsterdam **35**, 596, 1932; R. TOLMAN, various papers in Proc. Nat. Acad. Sci. and Phys. Rev.

[2]) I do not use "cosmic time" in the fantastic sense used by EDDINGTON in his Presidential Address to the Physical Society of London, 1931, when he pictured the universe (in his "cosmic time") as scurrying to evanescence.

by de SITTER) do not abolish the fact that a cosmic time is being employed; they merely correspond to the same observer using different watches. It is perfectly true that the value of ds^2 in the time-space of EINSTEIN or of FRIEDMANN-LEMAÎTRE is unchanged by a *change of co-ordinates.* What we are concerned with is a change of observer, and for a change of observer although the *value* of ds^2 is preserved the *form* of ds^2 is not preserved. For example if we change to an observer in uniform motion with respect to the first observer t will be transformed into a function of t' and the spatial co-ordinates in the new frame, and the metric will no longer be of the simple form (1) with a detached spatial part involving an $R(t)$ as coefficient. Cosmic time is the time kept by the observer for which the metric is precisely of the form (1) and no other. For isotropic curved space change of position of the observer leaves $[R(t)]^2 d\sigma^2$ unaffected, but changes of motion will cause a recognisable change in the *form* of ds^2. Thus this type of metric selects a special observer, and if he is taken to be "at rest" this is equivalent to re-introducing the idea of absolute rest, or absolute velocity, or absolute time which was destroyed by those experiments on which the special theory of relativity was founded. The theory I develop later necessarily involves an observer, but a structure is assigned to the world (by the extension of the principle of relativity) by which the observed *form* of the world — its distribution of matter and motion — is the same for all observers. It is then shown to be impossible to pick out a special observer relative to whom the complete system is at rest; the system appears to be at rest equally to all observers, whatever their velocity. Thus the notion of "at rest in the world" disappears, and with it " cosmic time". This result is achieved by the employment, not only of the differential invariant ds^2 (which only describes the geometry of a conceptual time-space consistent with constant light-velocity) but also of two integral invariants, which involve the *integrated* co-ordinates x, y, z, t and the velocities u, v, w of any particle of the system, measured by the observer, in his frame, from his natural origins of time and space.

A generalisation of the expanding-space theory which met the above criticisms would have to seek a metric ds^2 not only whose *value* but also whose *form* was invariant for all observers whatever their relative velocity.

24. But quite apart from the difficulty of accounting for the algebraic sign of the dilatation, and quite apart from the relativistic difficulty involved in introducing "cosmic time" the theory of "expanding space" encounters a further group of difficulties which I have never seen analysed. These are connected with the number of spatial dimensions required by

physics. Whatever the geometrical properties of the space adopted to describe the experimental results we call physics, physical phenomena proclaim in no uncertain way that *three* spatial dimensions are required for the observed phenomena to be possible. The arguments have been admirably summarised by Ehrenfest [1]) and similar arguments, connected with work by Volterra, have been adduced by Larmor [2]). For example, radiation as we know it seems to be possible only in space of 3 dimensions yet an *expanding* curved space necessarily implies that point-events and such phenomena as light propagation take place in a continuum of more than three spatial dimensions.

To make my meaning clear I propose to examine the state of affairs one dimension lower. It is to be understood that I am not merely arguing by analogy; I simply use the forms of expression which naturally come to hand one dimension lower because no expressive terminology equally familiar is available for describing expanding 3-space.

The expanding spherical world of Friedmann and Lemaître has often been likened to an expanding spherical balloon of which the surface represents "space" to two-dimensional inhabitants (observers) existing on it and on the fabric of which certain spots represent nebulae. The whole image is of course subtly misleading for there is nothing in this world of ours corresponding to the fabric constituting the balloon's surface. "Space" is no entity. To make the image perfect we must imagine the fabric of the balloon destroyed, leaving only the spots. The spots continue to lie, at any one instant [3]) in "cosmic time", on the geometrical locus defined by the sphere of radius determined by $R(t)$. Observations at any one "instant" are confined to, and events only occur in, this geometrical locus. The locus itself expands. Now the two-dimensional observers, being mathematicians and astronomers, would have already constructed for themselves Euclidean-space — "non spherical" space as they would call it — although it would contain points not ordinarily observable. For ordinary purposes they would use two-dimensional spherical space, but they would be aware that this space could be "immersed" in a 3-dimensional Euclidean space, of which at any one "instant" they could only observe a two-dimensional spherical section, namely the events occurring on this section. As time ("cosmic time") went on, however, and the space expanded, all points of the con-

[1]) Ann. d. Phys. **61**, 440, 1920.

[2]) C. R. du Congrès Intern. des Math. "Questions in Physical Interdetermination" 1930, p. 13.

[3]) This is not of course a world-wide instant save for the observer "at rest". It will be noticed that we are driven to use "cosmic time" in order even to describe expanding space in general terms.

structed Euclidean 3-space would in turn be the seat of events. The inhabitant-observers would find that quite apart from the different time-scales that could be used to describe the temporal relationships of events on the sphere, a third spatial co-ordinate (a function of one particular t) would be required to describe the observable events. A third spatial dimension would in fact be observable in the strict sense of the word, though at any one world-wide instant events would only be found on a 2-spread in this 3-space. Tracks of material particles, and paths of light rays, would occur in the 3-space — three co-ordinates x, y, z could in fact be assigned to any event and an (x, y, z) locus to a light track. But the observers would then be faced with the insoluble philosophical problem: "why is the whole of 3-space the seat of events, yet only a spherical section of it the seat of events at any one time?" They would find it necessary to attribute to each point of their 3-space properties which were a discontinuous function their cosmical time t, namely the property that for each point P of their 3-space there could be found a cosmical instant t', such that, for $t \neq t'$, P could not be the seat of an event but for $t = t'$ it could be the seat of an event. Further these physical properties of P and their discovery of a "real existence" for the third dimension would be in conflict with an ascertained result of *their* physics, to wit that physical occurrences of the character observed were compatible with two and only two spatial dimensions.

The argument applies as it stands to expanding spherical space of the type

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = [\Phi(t)]^2,$$

the expanding space of Friedmann and Lemaître. Hence the introduction of an expanding space is equivalent to the assertion that events can and do occur in a fourth spatial dimension but for each point P in the 4-space at one and only one instant t' in cosmic time. In addition to this difficulty we have the contradiction with the 3-dimensionality of space-relationships found in physics. If events occur in "expanding space", no 3-dimensional space can be found, for example, which contains the track of a light-ray, contradicting the fact that light as we know it[1]) can only be propagated in a 3-dimensional continuum. This is not the place to press these difficulties and a deeper discussion is required, but the foregoing considerations are sufficient to show that we cannot accept expanding spherical space simply because it is a possible solution of certain equations.

[1]) Obeying Huyghens's Principle.

25. Expanding *Euclidean* space has been considered recently by Einstein and de Sitter[1]). This involves no difficulties concerning dimensionality conditions, but it involves the difficulty that it seems to mean nothing in particular. The spatial metric

$$d\sigma^2 = -[R(t)]^2 (dx^2 + dy^2 + dz^2)$$

used by Einstein and de Sitter is simply equivalent to a change of length-scale with the time. We are accustomed in dynamics to the use of moving co-ordinate systems; the above metric seems to me to be simply a system of changing length-units. In this metric it can be shown that a point of fixed co-ordinates gives a red-shift at the origin which corresponds exactly to the rate of increase of the geodesic interval-distance from the origin to the point. In other words we may just as well use a "fixed space" and a "moving system" of "points" as a "non-static space" and a "co-ordinate-stationary" system of points. In the former case we have to explain why distant objects *recede*; in the latter case we have to explain why distant objects remain *the same size*. The concept of a non-static flat space is simply equivalent to a nest of co-ordinate boxes which expands with the time, expanding through ordinary Euclidean space. Any system of physical occurrences taking place in the non-static space can be described equally simply and with more insight as taking place in the fixed space, and the highly sophisticated notion of non-static space can be replaced without loss by the more homely fixed space[2]).

26. We now consider briefly the historic series of investigations which led to the present chaotic state of the "expanding space" theory. It is necessary to examine the arguments which have been used to justify the adoption of a curved, unclosed expanding space for the universe. The fundamental paper is Einstein's paper of 1917[3]) in which he candidly stated the difficulties with which he was faced at the time and submitted three main considerations which, he claimed, compelled him (a) to modify the law of gravitation previously used (b) to suggest a spherical form for the world. Einstein pointed out the following:

(1) An infinite world of Euclidean space filled with matter at a uniform density attracting according to the Newtonian law gives an infinite value

[1]) Proc. Nat. Acad. Sci. **18**, 213, 1932.

[2]) Prof. Freundlich has kindly pointed out to me that similarly expanding spherical space can be regarded as the continuous change of the length-unit on the surface of a fixed "sphere" But, just as with expanding flat space, this is merely equivalent to adopting a complicated way of describing a simple phenomenon.

[3]) Sitzungsber. d. Preuß. Akad. d. Wiss. 1917, S. 150.

for the gravitational potential at infinity and an infinite value also for the gravitational acceleration at infinity. The same result holds roughly for any pattern of gravitating mass-inequalities filling infinite space. To have a finite gravitational potential at infinity[1]) it appeared to EINSTEIN to be necessary, if Newtonian space were to be retained, to have a region of concentration such that the density decreased in all directions away from this point. The Newtonian universe would then be, as EINSTEIN elsewhere[2]) described it "a finite island in the infinite ocean of space". He went on: "this conception is not in itself satisfactory"; and it appeared to contradict the spirit of the principle of relativity by apparently picking out a special position in space and assigning to it special characteristics in relation to the world.

(2) A steady state of stellar velocities according to the BOLTZMANN law, extending to infinity, is incompatible with zero density at infinity, for finite gravitational potential at infinity. For when the difference of gravitational potentials is finite the ratio of the densities must be finite by BOLTZMANN's formula, and a density $\varrho_\infty = 0$ would therefore imply a central density $\varrho_c = 0$. EINSTEIN considered this difficulty scarcely explicable within the walls of the Newtonian theory.

(3) "We know from experience that for a suitably chosen co-ordinate system the velocities of the stars are small compared with the velocity of transmission of light[3])."

EINSTEIN's method of removing the supposed difficulties (1) and (2) was to assume a spherical (and therefore finite but unbounded) 3-space, the "surface" of $x_1^2 + x_2^2 + x_3^2 + x_4^2 = R^2$ and to fill it with matter at a uniform density. Consideration (3) then led him to attribute a zero velocity to this density-distribution as a first approximation. EINSTEIN then found that his gravitational equations

$$G_{\mu\nu} = -\varkappa\,(T_{\mu\nu} - \tfrac{1}{2}\,g_{\mu\nu}\,T)$$

were not fulfilled by *static* solutions with this space and this distribution of density and velocity. But if he added a term $-\lambda g_{\mu\nu}$ to the left hand side of the above equation where λ was a constant the equation as revised was found to be satisfied if only $\lambda = 1/R^2$. λ was called the "cosmical

[1]) A finite total mass is not necessary.
[2]) "The theory of relativity" (Eng. Trans.), 1921, Pt. III.
[3]) A. EINSTEIN, l. c.

constant"; and EINSTEIN suggested the equations $G_{\mu\nu} - \lambda g_{\mu\nu} = 0$ as the correct form of the gravitational equations in free space[1]).

Thus EINSTEIN's introduction of a curved finite unbounded space for the world and his modification of his original law of gravitation rested on definitely-stated physical considerations. That the curvature should be constant followed from EINSTEIN's postulate that all points in the world should be equivalent[2]).

At the same time in introducing the notion of matter "at rest" EINSTEIN abandoned the principal gain of his special theory of relativity, which was the denial of the existence of a meaning for "absolute rest". How do these three physical considerations stand to-day?

27. In the first place the velocities of the spiral nebulae, observed or about to be observed[3]) are not now all small compared with that of light. Hence the attribution of a velocity zero to the density-distribution is not justified. But further, since the velocities of the spiral nebulae are comparable with that of light, when we ourselves are observers, their velocities will also be comparable with that of light for any other observer travelling with any arbitrary velocity with respect to ourselves; for by the special theory of relativity, velocities comparable with light yield also velocities comparable with light when combined with any other velocity (law of combination of velocities). Hence not only is it now impossible to assign small velocities to the relevant celestial objects, but it is impossible also to pick out a preferential frame in which to reckon the actual velocities. Not only is the concept of "absolute rest in the world" philosophically unattractive, but it is physically impossible to decide on a frame having this property or to determine it by observation.

Considerations to be advanced later, of a very general character, result in the assigning to spiral nebulae of velocities ranging up to that of light in any Galileian frame whatever and so satisfactorily replace EINSTEIN's use of consideration (3).

28. With regard to consideration (2), the observed expansion phenomenon shows that the world of nebulae is not in a steady state. Thus BOLTZMANN's formula is inapplicable. The density- and velocity-distributions we shall later obtain differ widely from a BOLTZMANN and a MAXWELL distribution

---

[1]) It may be emphasized that the objections I have brought above against *expanding* spherical space have no force in connection with EINSTEIN's static spherical space. The objection to EINSTEIN's world is that involves the notion of absolute rest.

[2]) See quotation in § 1, para 1, above.

[3]) See § 2, para 10, above.

respectively, and are obtained without postulating either that the world is or is not in a steady state.

29. With regard to consideration (1) EINSTEIN believed that the postulate that all points in the world must be equivalent implied a constant density throughout the universe in the Newtonian frame of the observer, which in turn led to the view that a Newtonian infinite space is an impossibility. The second deduction is a consequence of the first but the first is incorrect. This can be seen as follows.

EINSTEIN's general theory of relativity implies only slight modifications in the Newtonian law, and hence provided we restrict ourselves to observers in uniform motion we can use to a first approximation the special theory of relativity. This is sufficient for my present purpose, which is to gain insight into the situation.

Consider therefore the distribution of matter and motion in the universe. We can make a map of this in the space-time of MINKOWSKI. We use space coordinates x, y, z and a time co-ordinate ict in the frame of an arbitrarily chosen observer. To each event corresponds a point P in this space-time, and to each material particle a world line. The whole distribution of matter and motion in the universe corresponds to a hyper-complex of curved world-lines. This hyper-complex is obtained by actually plotting the paths of the particles. Particles in motion in a given neighbourhood (x, y, z, ict) with different velocities correspond to world lines with different directions, the direction-cosines with respect to the axes chosen corresponding to the different velocities measured in the same frame. Different observers with different velocities correspond to different choices of axes, the direction of the t-axis defining the observer's velocity.

Consider now the whole hyper-complex of curves. Either:

(a) the system of world-lines is entirely chaotic or is a repeating pattern (crystalline structure) or

(b) the foregoing proposition is false.

Consider the two alternatives in turn.

(a) Any observer's instantaneous view of the world is the system of points given by a section of the complex by a plane $t =$ constant drawn through the observer's position (x, y, z, ict). If (a) holds, then the section will also obey (a) that is to say, the system of points will have either uniform density, or an absolutely irregular distribution (which is equivalent) or it will show a repeating pattern. In each case EINSTEIN's consideration (1) above shows that the system would have infinite gravitational potential at infinity and (a) is therefore ruled out. It cannot represent a real world.

(b) If (a) does not obtain, then the hyper-complex must possess some properties of symmetry. It might for example possess a sort of gradient in some definite direction; but this would pick out preferential directions in space-time which are denied by the special theory of relativity. Symmetry about a hyper-plane, a plane or a line or any number of these would pick out preferential directions in space-time and must be ruled out as contradicting observation. There remains only spherical symmetry about a point, say some point O.

Consider then a hyper-complex of world-lines having spherical symmetry about a given point O in space-time. Then all directions through O are equivalent. Take any point P, join it to O and describe the hyper-plane perpendicular to OP. This meets the hyper-complex in a 3-dimensional distribution of points which necessarily has spherical symmetry (in three dimensions) round P. This three-dimensional system is the view of the world by the observer whose world-line is OP, at the instant corresponding to the point P. Successive instants t in the history of the observer A whose world line is OP correspond to successive hyper-planes $t =$ constant parallel to the hyper-plane already described. They intersect the hyper-complex in a succession of 3-dimensional views of the universe which are always centred round A but change with the time t. For though the hyper-complex has the same local structure at all points P at a given "distance" r from O, different local structures will correspond to different distances r. Thus the observer represented by OP necessarily sees a world changing in time.

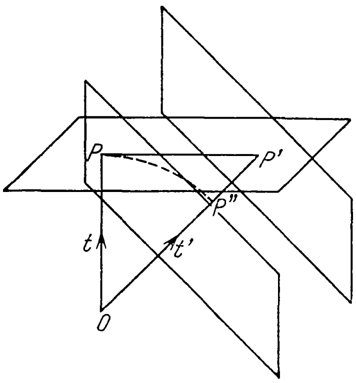

Fig. 2.

Now consider the view of the world obtained by another observer B, eccentric to A in A's view of the world. At the instant t in A's reckoning let P' be the point representing B. If B uses the same time-scale as A, i. e. if B is at rest relative to A, B's world-line will be a parallel to OP through P'; B will see himself as eccentric to A, but he will have the same view of the world as A. He can by astronomical observations determine the position of the centre (namely A at the instant t) and moreover by observation of the flux-phenomenon[1]) in the universe he will be able to determine the value of t (in A's reckoning) reckoned from the natural

[1]) I have not here proved that the flux phenomenon is an expansion phenomenon, but it is easy to see on general grounds that this is the only possibility.

origin O, namely OP. He can therefore determine the direction OP'. Now let him adopt OP' for his own time-scale, by suitable choice of his velocity relative to A. All he has to do is to set himself in motion with such a velocity V that it would bring him from P to his present position in the time t, i. e. a velocity numerically equal to PP'/t. This velocity corresponds precisely to the direction OP'. His view of the world is now the section of the hyper-complex by a hyper-plane perpendicular to OP'. From the symmetry of the hyper complex about O, this 3-dimensional system will have spherical symmetry about B's position namely P'. B at P' will not see the same view of the universe as A sees from P. But he will have already had A's P-view of the universe when he (B) was at P'', where P'' is a point on OP' such that $OP'' = OP$; for by the spherical symmetry of the world-lines about O, the hyper-plane through P'' perpendicular to OP'' intersects the hyper-complex in a 3-dimensional system of points of precisly the same form as the hyper-plane through P perpendicular to OP. Thus the *appearance* of the world to B at P'' is identical with the *appearance* of the world to A at P, although different particles play the same rôles in the two views, and the same particle plays different rôles. It follows that the complete succession of views of the universe experienced by A as he passes from $-\infty$ to $+\infty$ along OP is physically indistinguishable from the complete succession of views of the universe experienced by B as he passes from $-\infty$ to $+\infty$ along $OP''P'$. Each sees himself for all time as the centre of the universe.

Hence *any* two observers, wherever they be, can adopt velocities relative to one another so that each sees himself as the centre of the universe for all time. EINSTEIN's postulate is completely satisfied. All points in the universe are equivalent. The universe is completely ego-centric at all points [1]).

But now the density of the universe need not be, and will in fact not be, constant in the view of any one observer. The world-lines (not being uniformly dense in space-time yet possessing spherical symmetry round O) must become either more intensely or less intensely distributed as we go away from O. A little geometrical consideration will show that they cannot become more intense [2]). Hence the density falls off as we move away from O, and accordingly the density of point-intersections of the hyper-plane

---

[1]) An approach towards a similar result was made by H. Weyl, Phil. Mag. **9**, 936, 1930. See also ROBERTSON, Phil. Mag. **5**, 835, 1928.

[2]) This refers to the statistical system here considered, and in this connection the statement is true. But for the hydrodynamical-flow system considered in § 13, the density increases outwards up to $r = ct$, then becomes zero.

through P with the complex falls off as we proceed away from P in the hyperplane; that is, the density observed by A at P using OP as a time-axis ultimately thins away in all directions from P.

Provided it thins away sufficiently rapidly we shall have a finite gravitational potential at infinity. Einstein's postulate is then completely satisfied without the assumption of a uniform density-distribution in 3-space.

It will be seen that we have established the existence of a natural origin of space-time *in the map*, and a natural origin of space and of time in the reckoning, of any one observer. It is easily shown that these natural origins of space and time coincide with those obtained on kinematical grounds in § 2. But there is no such thing as a preferential region of space for there is no such thing as space without an observer. Each properly oriented observer will see every other properly oriented observer as running away from him, and each of them, equally legitimately, will regard himself as the "centre of space" (in his own reckoning) and the centre of the system of observed particles.

We have not found it necessary to "postulate" a "beginning of things". We have simply outlined from very general considerations, the only possible world which could exist consistently with Newtonian gravitation and with the special theory of relativity, and we have found that it satisfies Einstein's postulate for the world without the density-distribution being a uniform one. To *every* observer, the universe of particles appears an "island in space" ultimately thinning away in all directions.

Like most men of science I instinctively distrust very general arguments of the above type. Accordingly in Part II I shall give a[1]) mathematical investigation independent of general arguments which determines the forms of the world-lines consistent with special relativity. The conclusions will be found to be identical with those above.

It may seem paradoxical that it should be possible for every observer to see himself as the centre of a universe in which the density is not uniform. But Einstein himself showed that if a light-pulse starts out from a point occupied by an observer A, and simultaneously an observer B sets out from A with any uniform velocity, then each will see himself for all time as the centre of a spreading spherical light-pulse. Einstein's result is equally paradoxical, but is today accepted. Einstein's main point was that we cannot discuss light without first introducing an observer. Similarly

[1]) It may be of interest to mention that the analysis of Part II was completed before the arguments of this section occured to me or before I had concluded that "expanding space" is meaningless.

I wish to stress the fact that we cannot discuss the distribution of density and velocity in the world without first introducing an observer.

I conclude that EINSTEIN's difficulty (1) is not one which requires a refined theory of gravitation or of space for its removal. It disappears entirely when we examine the matter in the light of the Lorentz transformation and MINKOWSKI space-time, and do not forget to put in the observer. Thus the three physical considerations adduced by EINSTEIN are shown either to have no force today or to yield to further analysis.

When the effect of a gravitational field is taken into account the problem could probably be tackled on similar lines if EINSTEIN's general theory of relativity is adopted. The planes $t =$ constant will become surfaces which intersect the system of world lines in systems of points, each system being an instantaneous view of the universe for one observer, and we must seek such a geometry of co-ordinate surfaces that the succession of world views of one observer is indistinguishable from the succession of world-views of any other observer whose world-line is suitably oriented with respect to the system under observation. Instead however of trying to solve this problem outright I show in Part II how far the simple MINKOWSKI space-time properties may be used to analyse the actual structure of the world-lines to be expected in the universe[1]).

30. I have stressed the considerations brought forward by EINSTEIN in his 1917 paper because in the whole literature on the subject of world-structure they are almost the only considerations of a physical character which are to be found anywhere[2]). EINSTEIN's followers have contented themselves with development of a mathematical character, with the exception of EDDINGTON, who introduced philosophical considerations in an attempt to justify curved space. The story of these investigations is well known. DE SITTER[3]) introduced a world spherical in space and time, but this world was empty. FRIEDMANN[4]) and later (independently) LEMAÎTRE[5]) discovered the solution of EINSTEIN's gravitational equations for spherical

[1]) I by no means wish to imply that EINSTEIN's general theory of relativity is in my opinion the only way of generalising my analysis so as to include gravitation. The final solution of § 13 holds good for the world independent of any particular theory of gravitation.

[2]) TOLMAN's papers, however, definitely introduce physical considerations. I have partly discussed TOLMAN's papers in an unsigned review in Nature, "World Geometry in its Time-Relations", vol. 126, p. 742, 1930.

[3]) M. N. R. A. S. **78**, 3, 1917.

[4]) ZS. f. Phys. **10**, 377, 1922.

[5]) Ann. Soc. Sc. de Bruxelles **47**, A. Reprinted (translated) in M. N. **91**, 483, 1931.

space in which the spatial part of the metric depends on the time. In his Rhodes lectures at Oxford in 1931 Einstein[1]) pointed out that once non-static solutions were admitted there was no reason to retain the cosmical constant λ, for it was originally introduced solely to give a static solution for spherical space; as I had the privilege of hearing Einstein remark later, "il n'y a plus de λ". It was then (in 1931) pointed out by Heckmann[2]) that non-static solutions having been admitted, the curvature need not be positive and moreover λ could be positive zero or negative. Einstein and de Sitter[3]) pointed out that from the data of observation neither the sign of λ nor the sign of the curvature can be derived, and suggested the possibility of representing the observed facts without introducing curvature at all. Heckmann and de Sitter exhaustively considered the different cases that arise (1) for a world filled respectively entirely with radiation or entirely with matter (2) for $\lambda = 0, \pm 1$, (3) for curvature positive, zero and negative.

It thus appears that the cosmological constant and the curvature of space have been successively treated with suspicion, and that in fact purely mathematical considerations lead nowhere other than chaos. But if, as de Sitter affirms, curvature and cosmological constant are essentially indeterminable without some specific assumption, surely it would seem best to follow Einstein's own principle of introducing into mathematical physics only what can be observed. If the curvature of space cannot be determined, if it is essentially unobservable, then it should be rejected. The time has come when we should remember William of Occam's maxim: "Entia non sunt multiplicanda praeter necessitatem."

31. It remains however to mention Eddington's philosophical argument for a curvature of space. Eddington argued that there must be at each point of space some fundamental length, fixed by the geometry of the space, against which the electron can measure itself in order to take up a definite size. This argument should properly of described as anthropological rather than philosophical, for it employs the notion of the electron "knowing" what radius to assume. The objections to this way of arguing are numerous. Even if we are justified in assigning a definite length to an electron[4]),

---

[1]) See Sitzungsber. d. Preuß. Akad. d. Wiss. **235**, 1931.

[2]) O. Heckmann, Veröffentl. d. Universitäts-Sternwarte zu Göttingen, Heft 17, 1931 and Heft 23, 1932.

[3]) l. c.

[4]) It is curious that Eddington should fix on the least secure and most doubted of the electron's characteristics to fix the radius of space; *e* and *m* are so much better known.

Again, if "space" had a radius of curvature, and was expanding, why should the electron in question not expand with it and so conceal the expansion?

it would seem absurd to fix this by reference to something essentially unobservable; for space itself is essentially unobservable. Again it is impossible even to define "absolute length"; if all lengths were continually changing in fixed proportions to one another, no observable difference would be introduced. Hence in accordance with EINSTEIN's principle of economy we must construct a physics which dispenses with the notion of an absolute length. All we have reason to believe is that all electrons are alike; hence even on the anthropological viewpoint all that is necessary is that each electron should equate itself to all other electrons; it is the ratio of each electron to each other electron that is conserved, not the ratio of the radius of the electron to some essentially unobservable "radius of curvature of space". If there were only one electron in the world it would probably fill the world. Further we are not entitled to discuss an unobserved electron and when we introduce an observer spatial relations are immediately instituted between the observer and the electron, e. g. their separation. These brief remarks must suffice to indicate why, in my opinion, EDDINGTON's consideration in no way compels the adoption of a radius of curvature for space.

§ 6. *Time, space and laws of nature.* 32. In the literature of relativity, space, time and even space-time are apparently regarded as real entities existing in the physical world. For example near gravitating particles (in ordinary general relativity, apart from the cosmological problem) "space" is held to be "really" curved. A similar point of view is found with the philosophers. Thus BERTRAND RUSSELL[1]) distinguishes between three kinds of space: (1) the abstract spaces of geometry, (2) the "real" space of the physical world, (3) the separate spaces or perspectives of the observers. In contrast to this view the investigations of Part II of this paper are based on the view that there is no such thing as "real space" in the physical world, that there is no such thing as preexisting space with definite geometrical properties, whether created by matter (as many relativists hold) or not. On the contrary it is held that the observer can either (a) select any one of the spaces of pure geometry presented to him by the pure mathematician, use it in order to describe the phenomena of nature and then infer the laws obeyed by natural phenomena in this space; or alternatively (b) posit beforehand the laws of nature he wishes to see obeyed and then determine the space in which, in consequence, he must embed the occurrences he describes. I do not wish to maintain that each choice is

---

[1]) E. g. "Mysticism and Logic and other essays" and elsewhere.

absolutely arbitrary, but simply that the problem: "of what nature is the space of the world?" is essentially indeterminate.

33. Space is simply a convenient map for chronicling the spatial relationships of events. Any scheme for constructing the map can be adopted, but it is necessary to adopt *some* map-scheme before these relationships can be described in full. For example in descriptive geometries it is necessary to adopt axioms, although neither the axioms not the relationships to be analysed are of a metrical character. When the spatial relationships are not constant, we describe the totality of relationships by introducing time, and we make a space-time map of the world. We can again use any map-scheme that is convenient. Once the map-scheme is selected it becomes a question of describing as simply as possible the spatio-temporal relationships of events and so of isolating laws of nature.

Conversely it is sometimes possible to posit, in non-metrical form, some of the laws of nature we wish to see obeyed, and the scheme of the space-time map required may then be fixed. Thus space-time on the one hand and laws of nature on the other can be regarded as two different complementary aspects of that reality which is the flux in the relationships between events. Neither has an objective existence *per se*; neither are *given* in the real world. They only become definite on the introduction of an observer who makes a definite act of selection.

34. The supreme examples of the two modes of choice are offered by the gravitational theories of NEWTON and EINSTEIN respectively. NEWTON posited Euclidean space and "Newtonian" time, and determined with all but complete accuracy the law of gravitation holding in these species of space and time. This is an example of alternative (a). EINSTEIN posited the laws of mechanics and gravitation, or part of them, in the form of the principle of equivalence, supplemented this with another condition and determined the space-time necessary for any given distribution of matter and motion. This illustrates alternative (b). The principle of equivalence is in fact a comprehensive statement concerning certain laws of nature, made in the first instance without reference to the space-time metric to be subsequently used. Whether in Newtonian space and time the law of gravitation is exactly NEWTON's is a matter for observation; whether, assuming the principle of equivalence to be strictly true, space in the neighbourhood of a point-mass is exactly that given by the SCHWARZSCHILD metric (as tested by freely moving bodies) is a matter for observation. But in neither case is the truth of the original premises something verified or disproved by the observation. The observer resolves the observed relationship into a

space-time frame and a set of laws of nature which are inter-related. Simultaneously he resolves his space-time frame into space and time. But just as space and time have no separate independent existence according to EINSTEIN's special theory of relativity, so EINSTEIN's general theory of relativity may be interpreted as showing that space-time itself and laws of nature have no separate existence.

This point of view seems to me to throw light on the history of the development of solutions to the cosmological problem. We have seen how in attempting to construct a world-scheme, EINSTEIN alternately altered his space and altered his law of nature; he arbitrarily chose spherical space to represent the world and modified his law of gravitation to suit, then when "expanding spaces" were found to be a *mathematical* possibility he returned to his original law of gravitation. We have seen how according to HECKMANN and DE SITTER the sign of λ is inter-related with the sign of the space-curvature R, which again illustrates the complementary nature of space (sign of R) and laws of nature (sign of λ). It thus appears that the difficulties in which the current theory of the expanding universe finds itself partly disappear when we recognise the arbitrariness of imposing a particular nature on space. "Physical space" is simply a selected "conceptual space".

35. For example, whether "space" is infinite or not is a meaningless question. Whether the world of events has a finite extension in any selected space is however a question capable of answer. If the world of events has a finite spatial extension we can adopt a spherical space and fill it with events; or we can adopt infinite Euclidean space and partly fill it with events; or again we can map a finite extension on an infinite extension, as the earth's surface may be mapped on MERCATOR's projection. If the world of events has an infinite spatial extension, it will be best to use an infinite map on which to plot it, but this is not necessary — we could choose a finite map containing singular points.

36. In Part II I arbitrarily select flat Euclidean space as the seat of events. Simultaneously I consider a time-space map of MINKOWSKI type. Both are to be considered as maps for describing phenomena and observations. Whether either is "real" is not a legitimate question.

But the fact that the spirals form an expanding system expanding with a velocity-distance-proportionality law combined with the simplicity and directness of the kinematic explanation, is strong evidence that the nebulae constitute an unenclosed system and that no restriction exists limiting the maximum separation of two nebulae. In other words "space" is infinite in the only sense in which we are entitled to use the phrase. It will appear later that if there actually was a "beginning" of things at an epoch $t = 0$ then the observed world is necessarily finite, namely a sphere of radius $r = ct$ measured from any observer. If there was no beginning of things then it would appear that observation can extend beyond $r = ct$, but whether this outer region will contain systems of the spiral type inside it can scarcely be said.

II. Relativistic treatment of the kinematic solution.

§ 7. The velocity-distribution. 37. Before going on to the main problem of the *spatio-velocity* distribution of a moving system subject only to the postulate of special relativity and the extended postulate, I propose to consider first the simpler problem of the *velocity-distribution* of a swarm, under the same two postulates, when it is already given that the particles are describing straight lines with uniform velocities. In § 8 we proceed without the latter restriction.

38. *Formation of the functional equation.* Consider an observer A at rest in a particular Euclidean frame. His field of observation is a system of particles moving with uniform velocities along straight lines. We wish to determine the velocity-distribution of this system such that it will appear the same to *any* two observers moving with uniform velocity with respect to one another. Each observer can count the number of particles lying within given limits of velocity measured in his own frame. If two observers compare their observations using the special theory of relativity to transform the velocities, they will of course find that their observations are consistent, i. e. they will reconstruct the same reality. What we are to demand however — and this is a special case of the extended principle of relativity — is that the *form* of the velocity-distribution shall be the same to the two observers. In other words each must have the same velocity-perspective of the system. This is necessary in order to conceal the existence of any preferential velocity-frame. For if different observers had, presented to them, different velocity-pictures, some would be simpler than others, and on grounds of simplicity we could pick out a simplest, which would correspond to a preferential frame which we could call rest. If the system of particles is to give insight into the structure of the world — and I particularly do not wish to identify my particles at this stage with spiral nebulae — it is accordance with the spirit of relativity and Einstein's postulate that no one observer's view should be preferred to that of any other, and hence that all observers' views should be the same.

Suppose then that we have a second observer B moving with velocity V with respect to the first observer A in the direction of the x-axis of A's frame. B is to use a Euclidean-frame with axes parallel to those of A's frame, in motion with velocity $(V, 0, 0)$ relative to A. Let each observer count the particles lying in various velocity-ranges, independently of the spatial positions of the particles. Suppose that A's count results in a velocity-distribution law

$$f(u, v, w)\, du\, dv\, dw \tag{1}$$

measured in his own frame; (1) is the number of particles observed to have velocity-components between u and $u + du$, v and $v + dv$, w and $w + dw$ respectively. Let the second observer, using his own scales of length and time, obtain in his own frame the velocity law

$$f_1(u, v, w)\, du\, dv\, dw. \tag{2}$$

Since the velocity-components of a particle are constant, the fact that B may be counting at different "times" from A is immaterial; the lack of a meaning for simultaneity is of no consequence in this problem.

A particle observed by A to have the velocity (u, v, w) will be observed by B to have the velocity (u', v', w') where by the Lorentz-transformation for velocities[1])

$$u' = \frac{u - V}{1 - uV/c^2}, \quad v' = \frac{v(1 - V^2/c^2)^{1/2}}{1 - uV/c^2}, \quad w' = \frac{w(1 - V^2/c^2)^{1/2}}{1 - uV/c^2}. \tag{3}$$

Accordingly the particles observed by A to lie in the range u to $u + du$, ..., ..., will be observed by B to lie in the range u' to $u' + du'$, ..., ..., where the element of velocity-volume $du\, dv\, dw$ is connected with $du'\, dv'\, dw'$ by the relation

$$du'\, dv'\, dw' = \frac{\partial(u', v', w')}{\partial(u, v, w)}\, du\, dv\, dw.$$

Using (3) this gives

$$du'\, dv'\, dw' = \frac{(1 - V^2/c^2)^2}{(1 - uV/c^2)^4}\, du\, dv\, dw. \tag{4}$$

Now the number of particles counted by A as lying inside $du\, dv\, dw$ must be equal to the number counted by B as being inside $du'\, dv'\, dw'$. Hence

$$f(u, v, w)\, du\, dv\, dw = f_1(u', u', w')\, du'\, dv'\, dw', \tag{5}$$

or using (3) and (4)

$$f(u,v,w) = f_1\left(\frac{u - V}{1 - uV/c^2}, \frac{v(1 - V^2/c^2)^{1/2}}{1 - uV/c^2}, \frac{w(1 - V^2/c^2)^{1/2}}{1 - uV/c^2}\right) \frac{(1 - V^2/c^2)^2}{(1 - uV/c^2)^4}. \tag{6}$$

---

1) See e. g. the treatises of LAUE or of CUNNINGHAM. The formulae are of course due to EINSTEIN.

If now the swarm of particles is to obey the extended principle of relativity as regards its velocity-statistics it is necessary and sufficient that

$$f \equiv f_1. \tag{7}$$

It should be clearly understood that the principle of relativity and the extended principle of relativity make totally independent assertions about f. The special theory of relativity gives simply (6). Identity (7) is an entirely independent piece of information, not included in the ordinary principle of relativity.

We have accordingly the functional equation

$$f(u, v, w) = f\left(\frac{u-V}{1-uV/c^2}, \frac{v(1-V^2/c^2)^{1/2}}{1-uV/c'}, \frac{w(1-V^2/c^2)^{1/2}}{1-uV^2c^2}\right)\frac{(1-V^2/c^2)^2}{(1-uV/c^2)^4}. \tag{8}$$

We have also two other functional equations obtained by taking B to move along the y- and z-axes of A respectively. The question is whether (8) and the two similar equations possess a solution f independent of V and invariant for all changes of rectangular axes in the frame of A.

39. *The one-dimensional problem.* In order to illustrate simply the method adopted later for solving (8) I propose to consider as a parenthesis the analogous one-dimensional problem. Consider a system of particles moving in a straight line with a velocity-distribution $f(u)\,du$ with respect to a frame, as observed by an observer at rest in this frame. Let a second observer move along the straight line with velocity V relative to the frame of the first. Then

$$u' = \frac{u - V}{1 - uV/c^2}, \quad (3'); \qquad du' = \frac{1 - V^2/c^2}{(1 - uV/c^2)^2}\,du. \tag{4'}$$

The principle of special relativity, combined with the extended principle. for this system, leads as above to

$$f(u) = f\left(\frac{u - V}{1 - uV/c^2}\right)\frac{1 - V^2/c^2}{(1 - uV/c^2)^2}. \tag{8'}$$

Since we seek a solution holding for all values of V, a necessary condition is that (8′) shall be satisfied for all *small* values of V. Expanding by TAYLOR's theorem as far as the first power of V we have

$$f(u) = \left[f(u) - V\left(1 - \frac{u^2}{c^2}\right)f'(u)\right]\left[1 + \frac{2uV}{c^2}\right].$$

Hence a necessary condition is

$$\frac{2u}{c^2}f(u) - \left(1 - \frac{u^2}{c^2}\right)f'(u) = 0. \tag{9'}$$

This is an ordinary differential equation of which the solution is

$$f(u) = \frac{A}{c\,(1 - u^2/c^2)}.$$

A being a constant of integration. We now test whether this solution holds for all values of V. By direct evaluation we find that

$$\frac{A}{1 - \frac{1}{c^2}\left(\frac{u - V}{1 - u\,V/c^2}\right)^2} \cdot \frac{1 - V^2/c^2}{(1 - u\,V/c^2)^2} = \frac{A}{1 - \frac{u^2}{c^2}}$$

identically in u and V, and thus to all orders of V. We have thus found as the solution of the one-dimensional problem the distribution-law

$$\frac{A\,du}{c\,(1 - u^2/c^2)}.$$

The two observers, and hence any two observers, whatever their relative velocity, will now see exactly the same velocity-distribution. Any given particle will have different velocities to the two observers but the complete velocity-picture will be identical for the two observers. The velocities must necessarily range up to c, but not beyond; the assembly up to c provides just enough velocity-representatives to ensure that for any other observer the velocities just range up to c and no further.

It is of course remarkable (at first sight) that (8′) should possess a solution correct to all orders of V. We shall see the significance of this later.

40. *Solution of the functional equation for the velocity-distribution.* We return to our equation (8) and apply a similar method. Expanding to the first power of V and equating the coefficient of V to zero we get

$$\frac{4u}{c^2} f - \left(1 - \frac{u^2}{c^2}\right)\frac{\partial f}{\partial u} + \frac{u\,v}{c^2}\,\frac{\partial f}{\partial v} + \frac{u\,w}{c^2}\,\frac{\partial f}{\partial w} = 0. \tag{9}$$

Similarly from the two functional equations corresponding to B's motion along the y- and z-axes we have

$$\frac{4v}{c^2} f - \left(1 - \frac{v^2}{c^2}\right)\frac{\partial f}{\partial v} + \frac{v\,u}{c^2}\,\frac{\partial f}{\partial u} + \frac{v\,w}{c^2}\,\frac{\partial f}{\partial w} = 0, \tag{10}$$

$$\frac{4w}{c^2} f - \left(1 - \frac{w^2}{c^2}\right)\frac{\partial f}{\partial w} + \frac{w\,u}{c^2}\,\frac{\partial f}{\partial u} + \frac{w\,v}{c^2}\,\frac{\partial f}{\partial v} = 0. \tag{11}$$

Equation (9), (10), (11) form a set of 3 first order partial differential equations in three independent variables u, v, w and one dependent variable f. To solve them we follow the standard procedure[1]). We first render them

[1]) E. g., HOHEISEL, „Partielle Differentialgleichungen" (Sammlung Göschen), 1928.

homogeneous in the partial differential coefficients by seeking integrals of the implicit type

$$F\,(f,\, u,\, v,\, w) = 0, \tag{12}$$

writing

$$\frac{\partial f}{\partial u} = -\frac{\partial F}{\partial u}\Big/\frac{\partial F}{\partial f}$$

etc., and treating f as an independent variable. We then get

$$E_1 F \equiv \left[-4\frac{uf}{c^2}\frac{\partial}{\partial f} - \left(1-\frac{u^2}{c^2}\right)\frac{\partial}{\partial u} + \frac{uv}{c^2}\frac{\partial}{\partial v} + \frac{uw}{c^2}\frac{\partial}{\partial}\frac{\partial}{\partial w}\right]F = 0 \tag{13}$$

with two similar equations

$$E_2 F = 0, \quad E_3 F = 0 \qquad (13')\ (13'')$$

where E_1, E_2, E_3 are linear differential operators. Equations (13) may also be written in the form

$$\frac{1}{u}\frac{\partial F}{\partial u} = \frac{1}{v}\frac{\partial F}{\partial v} = \frac{1}{w}\frac{\partial F}{\partial w} = -\frac{4}{c^2}f\frac{\partial F}{\partial f} + \sum\frac{u}{c^2}\frac{\partial F}{\partial u} \tag{14}$$

Now by actual differentiation we find that

$$(E_1 E_2 - E_2 E_1)\,F \equiv \frac{1}{c^2}\left(-v\frac{\partial F}{\partial u} + u\frac{\partial F}{\partial v}\right) = 0 \tag{15}$$

and two similar equations. Since these equations (15) are already satisfied in virtue of the original equations, by (14), it follows by the well-known theory of systems of partial differential equations that the original equations (13), (13′), (13″) form a complete (vollständig) set. Since in this set the number of independent variables is four, namely u, v, w, f and the number of equations is three, it follows that the general integral F is a function of (4—3) or *one* function of f, u, v, w which is itself any integral of the equations. It is readily found that the simplest such integral is

$$(c^2 - u^2 - v^2 - w^2)^2 f \tag{16}$$

whence the general integral of (13), (13′), (13″) is

$$F\,(u,\, v,\, w,\, f) \equiv F\left((c^2 - u^2 - v^2 - w^2)^2 f\right).$$

F being arbitrary. Accordingly by (12) the solution of (9), (10), (11) is

$$f = \frac{A}{c^3\left(1 - \dfrac{u^2+v^2+w^2}{c^2}\right)^2} \tag{17}$$

where A is an arbitrary constant of zero dimensions[1]).

---

[1]) Note the square in the dominator, rendering the solution of the 3-dimensional problem significantly different from that of the 1-dimensional problem.

The solution (17) is accordingly a solution of the functional equation (8) correct to the first order in V. We now find that it is an exact solution for any value of V. For we have

$$\frac{A}{\left[1-\frac{1}{c^2}\,\frac{(u-V)^2+(v^2+w^2)(1-V^2/c^2)}{(1-u\,V/c^2)^2}\right]^2}\cdot\frac{(1-V^2/c^2)^2}{\left(1-\frac{u\,V}{c^2}\right)^4}$$

$$\equiv\frac{A}{\left[1-\frac{u^2+v^2+w^2}{c^2}\right]^2}$$

dentically in u, v, w, V. Moreover (17) is invariant for any change of rectangular axes. Accordingly (14) satisfies all the conditions required, and by its mode of establishment it is the only solution. It is therefore the answer to our problem.

41. It can now be seen why the functional equation (6) with $f_1 \equiv f$ possesses an exact solution independent of V. For by (4) $du\,dv\,dw$ is a covariant for a Lorentz transformation, and similarly $1-(u^2+v^2+w^2)/c^2$ is a covariant, for we have by direct substitution

$$\left(1-\frac{\Sigma\,u'^2}{c^2}\right)=\left(1-\frac{\Sigma\,u^2}{c^2}\right)\frac{1-V^2/c^2}{(1-u\,V/c^2)^2}.$$

The factor of covariance for $1-\Sigma u^2/c^2$ being the square root of the factor of covariance for $du\,dv\,dw$, it follows that a solution exists of the form (17); $du\,dv\,dw/(1-\Sigma u^2/c^2)^2$ is a Lorentz invariant.

It can now be seen without calculation that (17) is the desired solution of the problem. For if we plot the world-lines of the particles in Minkowski space, it is clear from the conditions of the problem that the number of world lines lying inside a solid angle (in 4-space) $d\Omega$ must be simply $A\,d\Omega$, integrated over the complete 4-space in x, y, z, t and by a formula to be established later, namely (28), $A\,d\Omega$ gives precisely the solution (17).

42. The velocity-distribution law

$$f(u, v, w)\,du\,dv\,dw=\frac{A\,du\,dv\,dw}{c^3\,(1-\sum u^2/c^2)^2} \tag{18}$$

has various interesting properties. It gives a continuous distribution of velocities for $|u, v, w| < c$, and this distribution is *complete* in the sense that the unclosed continuum of velocities so defined is necessary to solve the problem; for velocities arbitrarily close to $|u, v, w| = c$ are required to give velocities arbitrarily close to this limit in any other frame, and no velocities greater than c are required. For the velocity $|u, v, w| = c$

itself the analysis breaks down, since we cannot describe a velocity-interval $du\,dv\,dw$ round this value. This case requires a separate analysis, which I postpone to Part III, but I may mention here that particles moving with velocity $|u, v, w| = c$ prove to have precisely the properties of light, and the distribution of this light in frequency (compatible with the extended principle of relativity), can be evaluated.

As $|u, v, w| \to c$ the number of particles with velocity (u, v, w) tends to infinity. The observer at rest in any particular frame thus sees particles moving with all velocities up to that of light. Any other observer in uniform relative motion with respect to the first observer sees the same velocity-distribution; the LORENTZ contraction and change of time-reckoning conspire to replace each velocity-representative in the first observer's picture by some other in the second observer's picture. Neither observer can distinguish by inspection whether he is "at rest" or in "uniform motion" with respect to the system; consequently neither of these terms have any meaning. The phrase "the mean motion of the system" has no meaning. It should be noted that we have not assumed that the observer is necessarily attached to one of the members of the system; the velocity is the velocity of a point of view. But clearly no difference is introduced if we restrict observers to represented motions.

43. The important point is that all velocities up to c are represented. I shall prove later that the "global" velocity-distribution of any system of moving particles, however they interact, subject to the special theory of relativity and the extended principle of relativity is of the form now under consideration, but that the two principles together exclude gravitation, i. e. imply zero interactions[1]). It is therefore worth while at this stage enquiring what is the probable effect of gravitation in modifying the velocity-distribution. We have already considered the answer to this question on general grounds in Part I, § 3, par. 14.

Suppose that "initially" a system of *gravitating* particles has the velocity-distribution (18), and that the spatial distribution is a continuous one.

The problem is really an n-body problem with n infinite, but we can gain insight into the solution by the type of argument used in para. 14. It was there shown that there will result a number of closed sub-systems,

---

[1]) This is badly phrased. They imply zero *accelerations*, but as we shall see in § 13 this circumstance is compatible with any kind of gravitational interaction whatever, for our final density-distribution.

each escaping from one another but the members of each sub-system continuing to be gravitationally connected. The eventually-escaping sub-systems, like the original particles, have a velocity-distribution ranging up to c, and moreover of the general type of (18). For the sequence of escape-velocities converging to c is infinite in number, and hence between any escape-velocity v_n, however close to c, and c itself there will always exist further escape velocities relative to the sub-system of which v_n is a member, and so further escaping sub-systems.

This eventual "clotting" of a system with velocity-distribution up to c is something essentially different from the gravitational instability discussed by JEANS[1]). JEANS considered a primeval nebula of infinite extent, at rest, and showed that gravitational instability would cause it to break up and clot. But he did not enquire what he meant by "rest". He began with a distribution of matter without motion; he assigned to his primeval nebula the attribute of being "at rest". We have abolished the possibility of attaching a meaning to "at rest" and if we identify the leading characteristics of our system of particles with those of a primeval nebula, we can say that we have assigned to our primordial system the only possible distribution of motion that is consistent with the non-existence of a meaning for "absolute rest". In our system, unlike JEANS', the velocities range up to c, and gravitational agglomerations are formed not for reasons of instability but owing to the possibility of a sequence of velocities of escape. We thus approach the heart of the matter — the fact that the solution to both the relativistic cosmological problem and the problem of cosmogony (the world's past history) are both bound up with the distribution in the world not of matter alone but of matter and motion. JEANS and other have concentrated on matter alone and left out motion.

44. Before proceeding to the main problem, the determination of the distribution of both matter and motion, it is of some importance to notice a generalisation of the problem of the velocity-distribution. Suppose that we no longer restrict the particles to uniform rectilinear motion, but allow them to act on one another in any way, including collisions. The velocity distribution may now depend on the time. The observer A will be able to count the total number of particles with velocities lying inside u to $u + du, \ldots, \ldots$, at any one instant t by his clock, *summed through the whole of space*[2]). The observer B will be able to do likewise. Let the first

[1]) "Problems of Cosmogony and Stellar Dynamics" 1919.

[2]) I. e. integrated with respect to x, y, z.

observer's count result in a distribution law $f(u, v, w, t)\, du\, dv\, dw$ the second in the law $f_1(u', v', w', t')\, du'\, dv'\, dw'$. We now have the difficulty, that the positions of the particles in A's view reckoned at the world-instant t will not correspond to a world-instant t' in B's reckoning. For a given position of a particle in A's view, at a given t, t' will be a function of the position of the particle as well as of t, by the Lorentz transformation. It follows that if we impose the postulate of the extended principle of relativity, t must disappear from the distribution formula; for otherwise the two observers could not have the same view of the world. Thus distribution law (18) holds even when the particles interact with another subject to the two postulates of relativity. We shall later show, however, that under these conditions the paths are necessarily straight lines.

But the same result would seem to follow even if we go beyond special relativity and impose only the postulate of the extended principle. The latter would seem to be necessary for the world whatever its physics. The former must be replaced by some analogue of the Lorentz transformation for the actual relative motions of observers in the presence of the actual distribution of matter and motion — which analogue as far as I am aware has never been worked out. But t' will still be a function of t and the space co-ordinates of the separate particles. It follows that the integrated velocity-distribution must still be independent of t[1]). [It will not of course be exactly of the form (18), for this depends on the detailed formulae of the Lorentz transformation, although it will presumably include velocities up to c_∞, the velocity of light at a large distance from gravitating matter.] Thus it would appear that there should exist a definite velocity-distribution function for the world, independent of epoch, unchanging with the passage of time, and valid in the presence of gravitation. If this is not in fact true, we should have to abandon the extension-postulate and admit the possibility of the existence of observers with special points of view, i. e. we should have in effect to abandon EINSTEIN's cosmological principle.

§ 8. *The spatio-velocity distribution.* 45. We proceed now to solve the main problem of the paper. Consider a system of particles in motion in any manner whatever and acting on one another in any manner whatever but without collisions. We are to consider ourselves as totally unacquainted with dynamics or with the laws of interaction of the particles. Our object is to arrive at the laws of physics governing the particles by a study

---

[1]) We could avoid this conclusion only if it were found that the generalized Lorentz transformation implied that the formulae connecting u', v', w' with u, v, w involved the space and time co-ordinates as well.

of their motions. We are however to suppose ourselves aware of the Lorentz transformation formulae, supposed derived by light-signalling experiments and involving only the postulate of the constancy of the velocity of light. The Lorentz formulae, which are purely kinematical, permit the observations of any observer to be translated into the observations of a second observer moving with constant relative velocity with respect to the first observer. The general course of the argument which follows is that we first ascertain the form of the spatio-velocity distribution implied by the postulate of the constancy of the velocity of light (restricted relativity) and the postulate of the extended principle of relativity (EINSTEIN's cosmological principle). We then impose the condition that this spatio-velocity distribution formula shall actually represent, *statistically,* a concourse of material objects; this is done by a modification of BOLTZMANN's gas equation. Lastly we impose the condition that the *individual* particles shall actually move in the way prescribed by the distribution formula. This corresponds to the sequence: *relativity, statistics, dynamics.* We finally determine both the distribution function itself and the paths of the particles implied by the two postulates together with the postulate of the permanent existence of material objects.

46. Let an observer A be at rest in his own frame of reference, at the origin $x = 0$, $y = 0$, $z = 0$ of a Cartesian system of axes. Let him be provided with a clock keeping Newtonian time t[1]). At each instant t let him count the number of particles, lying inside the spatial range x to $x + dx, \ldots, \ldots,$ and the velocity range u to $u + du, \ldots, \ldots$. Let the number of particles so counted be

$$f(x, y, z, t, u, v, w)\, dx\, dy\, dz\, du\, dv\, dw. \tag{19}$$

Let a second observer B, carrying a clock synchronised with that of the first observer, leave the first observer at an arbitrary instant $t = 0$ (reckoned by either clock), in the direction of the first observer's x-axis, with velocity V. The observer B is to adopt a moving Cartesian system of axes $O' x' y' z'$ oriented parallel to those of A and to remain himself at the origin $x' = 0$, $y' = 0$, $z' = 0$. The instant $t = 0$ is not yet to be identified either with the natural kinematical origin of time of[2]) § 3 or with the time-position of the space-time origin of § 4, para 29; later it will be shown to coincide with both.

[1]) It is a consequence of § 13 that each observer can take his own time-sequence of world-pictures as constituting his clock.

[2]) As our system is not now necessarily one of uniform rectilineal motions, we do not in fact know that it possesses a kinematic natural origin of time.

Let the second observer also count the particles, using his own scales of length and time, and arrive at the distribution-law

$$f_1(x, y, z, t, u, v, w)\, dx\, dy\, dz\, du\, dv\, dw. \tag{20}$$

Any particular particle, at (x, y, z) at time t in A's reckoning, observed to possess the velocity u, v, w, will be observed by B to be at $x' y' z'$ at time t' in B's reckoning, with a velocity u', v', w', where by the Lorentz transformation

$$\left.\begin{aligned} x' &= \frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - V x/c^2}{(1 - V^2/c^2)^{1/2}} \\ u' &= \frac{u - V}{1 - uV/c^2}, \quad v' = \frac{v(1 - V^2/c^2)^{1/2}}{1 - uV/c^2}, \quad w' = \frac{w(1 - V^2/c^2)^{1/2}}{1 - uV/c^2} \end{aligned}\right\} \tag{21}$$

We cannot however say that all the particles observed by A inside $dx\, dy\, dz\, du\, dv\, dw$ at the instant t are observed by B inside $dx'\, dy'\, dz'\, du'\, dv'\, dw'$ at the instant t'; for the same world-wide instant corresponds to different values of t', each point $x' y' z'$ inside $dx'\, dy'\, dz'$ having its own t'[1]. We must therefore adopt a special procedure.

47. *Definition of intensity of world-lines.* Let observer A make a map of his observations in a conceptual space-time, by adopting any suitable set of co-ordinate surfaces $x = \text{const.}, \ldots, t = \text{const.}$, and plotting the observed world-line of each particle, one for each particle. If the distribution of matter and motion in the system is continuous, there will be one world-line through each "point" x, y, z, t, in each "direction" u, v, w. The complete aggregate of world-lines will therefore form a system in some ways analogous, in four dimensions, to a system of rays in a radiation-field in three dimensions.

But the analogy is not perfect. The intensity of radiation in a radiation field in a given direction can be defined as the number of photons (of given frequency range ν to $\nu + d\nu$) crossing an element of area dS perpendicular to the given direction, in time dt, inside a solid angle $d\omega$, divided by $dt\, dS\, d\omega$. The world-lines however form a set of geometrical curves in the chosen space-time and there is no analogy with the dt occurring in radiation theory. Instead, we define the intensity I of world-lines at a given point x, y, z, t in space-time in a given direction u, v, w as such that the number of world-lines crossing an elementary 3-way cross-section

[1]) The position is very similar to the calculation of the retarded potential for an electron in motion (WIECHERTS's formula).

$d\Sigma$ "perpendicular" to the given direction inside an elementary "cone" $d\Omega$ (in 4-space) is

$$I\, d\Sigma\, d\Omega.$$

48. Let us now take the observations of the second observer B and plot them in the same space-time by means of transformations (21). B must thus necessarily re-construct, identically, the same system of world-lines, for he is observing the same system. He will therefore arrive at the same value for the intensity I of world-lines in a given direction provided the two observers have the same rule for drawing the element $d\Sigma$ perpendicular to the given direction and the same methods of estimating the measures of $d\Sigma$ and $d\Omega$.

49. *Introduction of a metric.* It is the fundamental deduction in special relativity that the value of the differential form

$$c^2\, dt^2 - dx^2 - dy^2 - dz^2$$

is equal to the value of

$$c^2\, dt'^2 - dx'^2 - dy'^2 - dz'^2$$

when (x, y, z, t) and (x', y', z', t') are connected by (21). If we choose the metric ds^2 of our hitherto-arbitrary space-time map to be equal to this invariant, then by well-known theory the condition of perpendicularity and measures of length, etc., hold in all co-ordinates. We therefore adopt

$$ds^2 = -dx^2 - dy^2 - dz^2 + c^2\, dt^2.$$

It is to be particularly noted that we associate nothing physical or metaphysical with this ds^2. It is merely a constructed invariant rendering certain mathematical manipulations possible. The invariance of ds^2 is an algebraic consequence of relations (21), which in turn are deducible from the principle of the observed constancy of the velocity of light combined with EINSTEIN's light-signalling experiments.

It now follows[1]) that

$$I\, d\Omega\, d\Sigma$$

is the same for all observers (and not only for our two special observers A and B who parted company at this instant $t = 0$).

50. *Calculation of $d\Sigma$ and $d\Omega$.* We must now determine $d\Sigma$ and $d\Omega$ in terms of observable differentials $dx\, dy\, dz$ and $du\, dv\, dw$.

[1]) This is a consequence of the postulate of special relativity only. Of course the observers see the same picture in *time-space*, merely because they are observing the same world. But that they will see the same picture *in space and in time,* i. e. through the eyes of their own special co-ordinate systems and time-scales, is only given by the second postulate of relativity (the generalisation of EINSTEIN's cosmological principle), first employed in § 52.

Given the existence of a metric, the whole apparatus of vector and tensor analysis follows. For simplicity and to make all the co-ordinates of the physical dimensions of a length I propose to take as co-ordinates $x_1 = x$, $x_2 = y$, $x_3 = z$, $x_4 = ct$, but I shall always write formulae as if transformed back into x, y, z, t. It is particularly to be noted that our space-time is only a convenient map of what is happening.

The observed velocity u, v, w defines a direction of direction-cosines

$$\frac{u}{(c^2 - \sum u^2)^{1/2}},\quad \frac{v}{(c^2 - \sum u^2)^{1/2}},\quad \frac{w}{(c^2 - \sum u^2)^{1/2}}\cdot\quad \frac{c}{(c^2 - \sum u^2)^{1/2}} \tag{22}$$

which is the direction of the normal to $d\Sigma$. The magnitude of $d\Sigma$ is the projection of the element of 3-volume $dx\,dy\,dz$ normal to this direction. All world-lines passing through $d\Sigma$ in the given direction through a given point intersect the surface $t =$ constant (in an arbitrarily chosen co-ordinate system), though the given point, in the element $dx\,dy\,dz$; $t =$ constant giving of course the world-wide instant for the observer who corresponds to the co-ordinate system chosen. It follows that

$$d\Sigma = \frac{c}{(c^2 - \sum u^2)^{1/2}}\, dx\,dy\,dz \tag{23}$$

for $c/(c^2 - \Sigma u^2)^{1/2}$ is the cosine of the angle between $d\Sigma$ and $dx\,dy\,dz$.

We have now to determine $d\Omega$ in terms of $du\,dv\,dw$. We let x, y, z, t denote for the moment *any* current co-ordinates, adopt a change of variables to $R, \vartheta, \varphi, \psi$, where

$$\left.\begin{aligned} x &= R \sinh\vartheta \sin\varphi \sin\psi, \\ y &= R \sinh\vartheta \sin\varphi \cos\psi, \\ z &= R \sinh\vartheta \cos\varphi, \\ ct &= R \cosh\vartheta, \end{aligned}\right\} \tag{24}$$

and consider the pseudo-hyper-sphere

$$c^2t^2 - x^2 - y^2 - z^2 = R^2. \tag{25}$$

We can define the 4-space solid angle $d\Omega$ in two ways. Either (1) we can equate the general element of 4-volume $dx\,dy\,dz\,c\,dt$ to $R^3 dR\,d\Omega$; or (2) we can take an element of the "surface" of (25), namely $R^3 d\Omega$, and equate its projection on the plane $t =$ constant, namely $R^3 d\Omega \cosh\vartheta$, to the 3-space cross-section $dx\,dy\,dz$[1]) on this plane. Now from (24)

$$\begin{aligned} dx\,dy\,dz\,c\,dt &= \frac{\partial(x, y, z, ct)}{\partial(R, \vartheta, \varphi, \psi)}\, dR\,d\vartheta\,d\varphi\,d\psi \\ &= R^3 \sinh^2\vartheta \sin\varphi\, dR\,d\vartheta\,d\varphi\,d\psi. \end{aligned}$$

[1]) This $dx\,dy\,dz$ is not to be confused with the $dx\,dy\,dz$ of (23) etc.

Using definition (1), this gives

$$d\Omega = \sinh^2\vartheta \sin\varphi \, d\vartheta \, d\varphi \, d\psi. \tag{26}$$

Again, for the projection of the element of the surface of (25) we have by (24)

$$dx\,dy\,dz^{1)} = \left[\frac{\partial(x, y, z)}{\partial(\vartheta, \varphi, \psi)}\right]_{R=\text{const}} d\vartheta\,d\varphi\,d\psi$$
$$= R^2 \cosh\vartheta \sinh^2\vartheta \sin\varphi \, d\vartheta \, d\varphi \, d\psi.$$

Using definition (2) this gives

$$d\Omega = \sinh^2\vartheta \sin\varphi \, d\vartheta \, d\varphi \, d\psi. \tag{26'}$$

Comparison of (26) and (26′) shows that the two definitions of $d\Omega$ are consistent.

We now have to express $du\,dv\,dw$ in terms of $d\vartheta\,d\varphi\,d\psi$. The above elementary "cone" $d\Omega$, given by (26), has its axis parallel to the direction $x:y:z:ct$ given by (24). We must therefore choose ϑ, φ, ψ so that the direction corresponding to the velocity $u, v, w,$ is parallel to the direction (24). This requires

$$\left.\begin{aligned} u &= c \tanh\vartheta \sin\varphi \sin\psi, \\ u &= c \tanh\vartheta \sin\varphi \cos\psi, \\ w &= c \tanh\vartheta \cos\varphi, \end{aligned}\right\} \tag{27}$$

for this gives the right direction-cosines by (22). Hence using (27)

$$du\,dv\,dw = \frac{\partial(u, v, w)}{\partial(\vartheta, \varphi, \psi)} d\vartheta\,d\varphi\,d\psi$$
$$= c^3 \operatorname{sech}^2\vartheta \tanh^2\vartheta \sin\varphi \, d\vartheta\,d\varphi\,d\psi.$$

Using (26) we have

$$du\,dv\,dw = c^3 \operatorname{sech}^4\vartheta \, d\Omega$$

or

$$d\Omega = \frac{du\,dv\,dw}{c^3[1 - \sum u^2/c^2]^2}. \tag{28)$^{2)}$}$$

51. *The distribution function f in terms of I.* Since the world-lines passing through $dx\,dy\,dz$ inside $d\Omega$ represent an equal number of visible particles inside $dx\,dy\,dz$ with velocities inside $du\,dv\,dw$, we must have

$$I\,d\Omega\,d\Sigma = f(x, y, z, t, u, v, w)\,dx\,dy\,dz\,du\,dv\,dw.$$

---

[1]) This $dx\,dy\,dz$ is not to be confused with the $dx\,dy\,dz$ of (23) etc.

[2]) I have given the derivation of (23) and (28) in some detail partly because I have not been able to find them in the literature, partly become I wish to show that I am using concrete mathematics and not loose geometrical analogies. Formula (28) gives great insight in to the velocity-distribution law found in § 7.

Hence by (23) and (28)

$$f(x, y, z, t, u, v, w) = \frac{I}{c^3 [1 - \Sigma u^2/c^2]^{5/2}}. \qquad (29)$$

But the same argument holds in any other frame. Hence

$$f_1(x', y', z', t', u', v', w') = \frac{I}{c^3 [1 - \Sigma u'^2/c^2]^{5/2}}. \qquad (29')$$

Now I is a definite number characteristic of the world-line structure in a given direction near a given point. I can be expressed as a function of x, y, z, t, u, v, w, or as a function of $x', y', z', t', u', v', w'$. Suppose then that

$$I = \varphi(x, y, z, t, u, v, w) = \varphi_1(x', y', z', t', u', v', w'). \qquad (30)$$

52. *Application of the extended principle of relativity.* Now let us apply to our two particular observers A, B, who parted company at $t = 0$, $t' = 0$ in their x-directions with speed V, the postulate of the extended principle of relativity. That is so say, let us impose the condition that each observer sees the same sequence of views of the world, independent of the relative speed V. Then f must be the same function of $x, \ldots t, \ldots w$ as f_1 is of $x', \ldots t', \ldots w'$ i. e.

$$f \equiv f_1. \qquad (31)$$

Hence by (29) and (29') we must have

$$\varphi \equiv \varphi_1. \qquad (32)$$

Hence, by (30),

$$\varphi(x, y, z, t, u, v, w) = \varphi(x' y', z', t', u', v', w'). \qquad (33)$$

That is,

$$\varphi(x, y, z, t, u, v, w) = \varphi\left(\frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, y, z, \frac{t - Vx/c^2}{(1 - V^2/c^2)^{1/2}}, \frac{u - V}{1 - uV/c^2}, \frac{v(1 - V^2/c^2)^{1/2}}{1 - uV/c^2}, \frac{w(1 - V^2/c^2)^{1/2}}{1 - uV/c^2}\right) \qquad (34)$$

identically in $x, \ldots t, \ldots w$ for all V.

This functional equation and two similar ones obtained by letting B move along the y- and z-axes respectively must be satisfied by the function φ, and in addition φ must be of the same form for all changes of spatial rectangular axes. It is to be noted that we cannot infer this functional equation from the special theory of relativity alone. By special relativity alone we can only infer that the *value* of φ is independent of axes. The extended postulate is required to infer that the *form* of φ is independent of axes. φ is not merely an invariant in the sense that the ds^2 of *general* relativity theory is an invariant; all that is posited about such ds^2 is that its numerical value is unaltered by change of co-ordinates. φ resembles the ds^2 of *special* relativity theory, whose *form* as well as whose value is conserved. But φ involves velocity-components as well as space-time coordinates. And it is an integral invariant, that is to say as far as we have gone at present it involves co-ordinates measured from special zeros.

53. *Solution of the functional equation.* There are two different ways of solving equation (34). The one method (a) is to construct the partial differential equations necessarily satisfied for V small, and solve them. The other method (b) is to examine the known invariants under Lorentz-transformations. I consider them in turn.

(a) Expanding the left-hand side of (34) by Taylor's theorem to the first power in V and equating the coefficient of V to zero we have, as in para (39),

$$D_1\,\varphi \equiv \left(1-\frac{u^2}{c^2}\right)\frac{\partial\varphi}{\partial u}-\frac{uv}{c^2}\frac{\partial\varphi}{\partial v}-\frac{uw}{c^2}\frac{\partial\varphi}{\partial w}+t\frac{\partial\varphi}{\partial x}+\frac{x}{c^2}\frac{\partial\varphi}{\partial t}=0. \quad (35)$$

Similarly by cyclical interchange of x, y, z and u, v, w we have

$$D_2\,\varphi \equiv \left(1-\frac{v^2}{c^2}\right)\frac{\partial\varphi}{\partial v}-\frac{vu}{c^2}\frac{\partial\varphi}{\partial u}-\frac{vw}{c^2}\frac{\partial\varphi}{\partial u}+t\frac{\partial\varphi}{\partial y}+\frac{y}{c^2}\frac{\partial\varphi}{\partial t}=0, \quad (35')$$

$$D_3\,\varphi \equiv \left(1-\frac{w^2}{c^2}\right)\frac{\partial\varphi}{\partial w}-\frac{wu}{c^2}\frac{\partial\varphi}{\partial u}-\frac{wv}{c^2}\frac{\partial\varphi}{\partial v}+t\frac{\partial\varphi}{\partial z}+\frac{z}{c^2}\frac{\partial\varphi}{\partial t}=0. \quad (35'')$$

These form a set of three linear homogeneous partial differential equations in seven independent variables and one dependent variable. We seek the general integral.

Applying a standard procedure we find new operators D_4, D_5, D_6 given by

$$\frac{1}{c^2}D_4\,\varphi \equiv (D_2 D_3 - D_3 D_2)\,\varphi$$

$$= \frac{1}{c^2}\left[-w\frac{\partial\varphi}{\partial v}+v\frac{\partial\varphi}{\partial w}+y\frac{\partial\varphi}{\partial z}-z\frac{\partial\varphi}{\partial y}\right]=0, \quad (36)$$

$$\frac{1}{c^2}D_5\,\varphi \equiv (D_3 D_1 - D_1 D_3)\,\varphi$$

$$= \frac{1}{c^2}\left[-u\frac{\partial\varphi}{\partial w}+w\frac{\partial\varphi}{\partial u}+z\frac{\partial\varphi}{\partial x}-x\frac{\partial\varphi}{\partial z}\right]=0, \quad (36')$$

$$\frac{1}{c^2}D_6\,\varphi \equiv (D_1 D_2 - D_2 D_1)\,\varphi$$

$$= \frac{1}{c^2}\left[-v\frac{\partial\varphi}{\partial u}+u\frac{\partial\varphi}{\partial v}+x\frac{\partial\varphi}{\partial y}-y\frac{\partial\varphi}{\partial x}\right]=0. \quad (36'')$$

This generates three new partial differential equations which φ must satisfy. Repeating the procedure we find that

$$(D_5 D_6 - D_6 D_5) \equiv - D_4,$$

and two similar identities; and in like manner

$$(D_1 D_4 - D_4 D_1) \equiv 0,$$
$$(D_1 D_5 - D_5 D_1) \equiv - D_3,$$
$$(D_1 D_6 - D_6 D_1) \equiv + D_2,$$

and six further similar identities. Since no new equations are generated by this process, the six equations (35), (35′), (35″), (36), (36′), (36″) form a complete (vollständig) set. They are not however linearly independent. Multiplying (36), (36′), (36″) in turn by x, y, z and adding, and then by u, v, w and adding, we find

$$\sum \frac{\partial \Phi}{\partial u}(v z - w y) = 0, \tag{37}$$

$$\sum \frac{\partial \Phi}{\partial x}(v z - w y) = 0. \tag{37'}$$

But (35), (35′) and (35″) may be written in the form

$$-\frac{u}{c^2}\left(\sum u \frac{\partial \Phi}{\partial u}\right) + \frac{\partial \Phi}{\partial u} + t \frac{\partial \Phi}{\partial x} + \frac{x}{c^2}\frac{\partial \Phi}{\partial t} = 0$$

and two similar equations. Multiplying these in turn by $vz - wy$, $wx - uz$, $uy - vx$ and adding we get

$$\Sigma (v z - w y) \frac{\partial \Phi}{\partial u} + t \, \Sigma (v z - w y) \frac{\partial \Phi}{\partial x} = 0$$

which is satisfied identically in virtue of (37) and (37′).

It follows that the complete set is equivalent to five equations, namely the group (36) and any two of the group (35). Since we have 5 equations and 7 variables, the general solution Φ is an arbitrary function of (7—5) or *two* functions of these variables, which are themselves independent integrals of the system. The integrals are found without much trouble to be

$$t^2 - \frac{x^2 + y^2 + z^2}{c^2} \tag{38}$$

and
$$\frac{\left[t-\frac{u\,x+v\,y+w\,z}{c^2}\right]^2}{1-\frac{u^2+v^2+w^2}{c^2}}. \tag{39}$$

It will be found convenient to write

$$X = t^2 - \frac{x^2+y^2+z^2}{c^2}, \tag{40}$$

$$Y = 1 - \frac{u^2+v^2+w^2}{c^2}, \tag{41}$$

$$Z = t - \frac{u\,x+v\,y+w\,z}{c^2}. \tag{42}$$

Then the general integral of the equations is

$$\varphi = \varphi\left(X, \frac{Z^2}{Y}\right). \tag{43}$$

Now $f\,dx\,dy\,dz\,du\,dv\,dw$ is a pure number, a number of particles. Hence f is of dimensions -3 in length and -3 in velocity. Y is of zero dimensions, and c is a velocity. Hence by (29) I and so φ is of dimensions -3 in length. Without loss of generality we can therefore write

$$\varphi = \frac{1}{c^3 X^{3/2}}\,\psi\left(X, \frac{Z^2}{Y}\right) \tag{44}$$

where ψ is of zero dimensions. It can only be a function of pure numbers. But X is of dimensions $+2$ in time, Z^2/Y of dimensions $+2$ in time. Hence ψ must be a function of Z^2/XY only, say

$$\psi \equiv \psi\,(Z^2/X\,Y)$$

and

$$\varphi = \frac{\psi\,(Z^2/X\,Y)}{c^3 X^{3/2}}. \tag{45}$$

We now notice that X is an invariant under a Lorentz transformation (as is well known),

$$X' = t'^2 - \frac{\Sigma\,x'^2}{c^2} = t^2 - \frac{\Sigma\,x^2}{c^2} = X, \tag{46}$$

and that Y and Z are covariants,

$$Y' = 1 - \frac{\Sigma\,u'^2}{c^2} = \left(1-\frac{\Sigma\,u^2}{c^2}\right)\frac{1-V^2/c^2}{(1-u\,V/c^2)^2} = Y\,\frac{1-V^2/c^2}{(1-u\,V/c^2)^2}, \tag{47}$$

$$Z' = t' - \frac{\Sigma\,u'\,x'}{c^2} = \left(t-\frac{\Sigma\,u\,x}{c^2}\right)\frac{(1-V^2/c^2)^{1/2}}{1-u\,V/c^2} = Z\,\frac{(1-V^2/c^2)^{1/2}}{1-u\,V/c^2}. \tag{48}$$

It follows that Z^2/Y is an invariant. Hence the solution (45) satisfies (34) not only for V small but for all values of V.

We have thus that the intensity-distribution of world-lines is given by

$$I\,d\Omega\,d\Sigma = \frac{\psi(Z^2/XY)}{c^3 X^{3/2}}\,d\Omega\,d\Sigma \tag{49}$$

and hence by (29) the distribution observed at the instant t is given by

$$f(x, y, z, t, u, v, w)\,dx\,dy\,dz\,du\,dv\,dw = \frac{\psi(Z^2/XY)}{c^6 X^{3/2} Y^{5/2}}\,dx\,dy\,dz\,du\,dv\,dw. \tag{50}$$

(b) A much shorter path to the same result is to notice that I is a function of position and direction in space-time, and that if the postulate of the extended principle is to be satisfied, I must be of the same form whatever the Cartesian frame chosen, provided the frame has an assigned origin $x = 0$, $y = 0$, $z = 0$, $t = 0$. I is therefore a 4-scalar. The only two vectors concerned are

$$x, \quad y, \quad z, \quad ct,$$

and

$$u/Y^{1/2}, \quad v/Y^{1/2}, \quad w/Y^{1/2}, \quad c/Y^{1/2}.$$

The only independent scalars associated with these are their absolute values and the cosine of the angle between them, which are respectively c^2X, 1 and $c^2Z/Y^{1/2}$ (the metric being $ds^2 = -dx^2 - dy^2 - dz^2 + c^2dt^2$). Hence $\mathrm{I} = \Psi(X, Z/Y^{1/2})$, which is the result already found.

54. *Fulfilment of the extended principle of relativity.* So far the postulate of the extended principle of relativity has been satisfied for any two observers who set off from one another with an arbitrary velocity V. In formula (50) the origin of x, y, z is the position of either observer (in his own frame) and the origin of t is the moment of separation of the two observers, reckoned in either frame. Each of these two observers sees unfolding before him exactly identical views of the universe.

Now consider any third observer C at any point of space. If he is at rest relative to A, he will see himself as eccentric to the distribution, for from the form of (50) it is clear that the space-distribution at any instant is spherically symmetrical about A and this spherical symmetry round A will be recognizable by C. But now let C assume such a velocity as would have brought him from the centre A to his present eccentric position C in time t where t is the epoch of the moment considered. Then he will immediately become an observer of type B, and he will see unfolding before him the same view of the universe as A sees, by the conditions satisfied by f;

in particular C will see himself the centre of the distribution. This velocity is experimentally ascertainable. It lies in the line AC, is of amount Ac/t ans is less than c provided $AC < ct$. Thus all points inside the sphere $r = ct$ are equivalent in the sense that an observer at *any* point can assume a velocity such that he then appears to be the centre of the distribution and from this view-point he will see unfolded exactly the same sequence of configurations as any other observer sees who chooses his velocity (i. e. his time-scale) so as to be himself the centre of the distribution.

The same is *algebraically* true for points outside the sphere $r = ct$, although the Lorentz transformation then involves imaginaries, since $V > c$. Points outside $r = ct$ are however essentially inaccessible in the time that has elapsed from $t = 0$.

Thus the extended principle of relativity is satisfied everywhere. It will be seen that we have been led by rigorous mathematical procedure to precisely the state of affairs sketched on general grounds in Part I, § 5, para. 29. In particular we note the emergence of spherical spatial symmetry and a natural time origin $t = 0$, with corresponding spherical symmetry in time-space. But any concept of absolute position in space, absolute velocity or absolute time has disappeared.

§ 9. The analogue of Boltzmann's equation. 55. *Principle of permanence.* The question arises whether we can determine the form of the function ψ. We have so far not made use of the fact that the distribution is to describe a system of permanent objects. We have to satisfy the condition that the distribution prescribed for time t shall actually change kinematically during the interval Δt into the distribution prescribed for time $t + \Delta t$. A similar condition is well-known in gas-theory as Boltzmann's equation[1]). We cannot however take over Boltzmann's equation as it stands for the usual form of this equation is obtained by taking a field of force which is a function of position only, whilst we are quite unaware of the nature of the interactions between our particles. These may be a function of velocity, for example.

The equations of motion of any particle of the system are

$$\frac{dx}{dt} = u, \quad \frac{dy}{dt} = v, \quad \frac{dz}{dt} = w,$$

$$\frac{du}{dt} = p, \quad \frac{dv}{dt} = q, \quad \frac{dw}{dt} = r,$$

[1]) See e. g. Jeans, Kinetic Theory of gases.

where p, q, r are the components of acceleration of the particle x, y, z at time t, when moving with velocity u, v, w. These are *descriptively* functions of the 7 independent variables $x, \ldots t, \ldots w$. Take a small time-interval Δt and write

$$t_1 = t + \Delta t$$

$$x_1 = x + u\,\Delta t, \qquad y_1 = y + v\,\Delta t, \qquad z_1 = z + w\,\Delta t$$

$$u_1 = u + p\,\Delta t, \qquad v_1 = v + q\,\Delta t, \qquad w_1 = w + r\,\Delta t.$$

Then the condition for a collection of permanent objects is

$$f(x, y, z, t, u, v, w)\, dx\, dy\, dz\, du\, dv\, dw$$

$$= f(x_1, y_1, z_1, t_1, u_1, v_1, w_1)\, dx_1\, dy_1\, dz_1\, du_1\, dv_1\, dw_1, \tag{51}$$

For the number of particles inside $dx \ldots dw$ at time t must equal the number of particles inside $dx_1 \ldots dw_1$, at time t_1.

Now we have

$$\left.\begin{aligned} dx_1\, dy_1\, dz_1\, du_1\, dv_1\, dw_1 &= \frac{\partial (x_1, y_1, z_1, u_1, v_1, w_1)}{\partial (x, y, z, u, v, w)}\, dx\, dy\, dz\, du\, dv\, dw \\ &= dx \ldots dw \begin{vmatrix} 1, & 0, & 0, & \Delta t, & 0, & 0 \\ 0, & 1, & 0, & 0, & \Delta t, & 0 \\ 0, & 0, & 1, & 0, & 0, & \Delta t \\ \frac{\partial p}{\partial x}\Delta t, & \frac{\partial p}{\partial y}\Delta t, & \frac{\partial p}{\partial z}\Delta t, & 1+\frac{\partial p}{\partial u}\Delta t, & \frac{\partial p}{\partial v}\Delta t, & \frac{\partial p}{\partial w}\Delta t \\ \frac{\partial q}{\partial x}\Delta t, & \frac{\partial q}{\partial y}\Delta t, & \frac{\partial q}{\partial z}\Delta t, & \frac{\partial q}{\partial u}\Delta t, & 1+\frac{\partial q}{\partial v}\Delta t, & \frac{\partial q}{\partial w}\Delta t \\ \frac{\partial r}{\partial x}\Delta t, & \frac{\partial r}{\partial y}\Delta t, & \frac{\partial r}{\partial z}\Delta t, & \frac{\partial r}{\partial u}\Delta t, & \frac{\partial r}{\partial v}\Delta t, & 1+\frac{\partial r}{\partial w}\Delta t \end{vmatrix} \\ &= \left[1 + \Delta t\left(\frac{\partial p}{\partial u} + \frac{\partial q}{\partial v} + \frac{\partial r}{\partial w}\right)\right] dx\, dy\, dz\, du\, dv\, dw. \end{aligned}\right\} \tag{5}$$

Hence

$$f(x, y, z, t, u, v, w)$$

$$= \left(1 + \Delta t\, \Sigma \frac{\partial p}{\partial u}\right) f(x + u\,\Delta t, \ldots, \ldots, t + \Delta t, \ldots, \ldots, \ldots w + r\,\Delta t).$$

Applying TAYLOR'S theorem we have accordingly

$$\frac{\partial f}{\partial t} + \Sigma u \frac{\partial f}{\partial x} + \Sigma p \frac{\partial f}{\partial u} + f \Sigma \frac{\partial p}{\partial u} = 0. \tag{53}$$

This may be conveniently written

$$\frac{1}{f} \Delta f + \Sigma \frac{\partial p}{\partial u} = 0, \tag{54}$$

where
$$\Lambda \equiv \frac{\partial}{\partial t} + \Sigma u \frac{\partial}{\partial x} + \Sigma p \frac{\partial}{\partial u} \cdot \tag{55}$$

Equation (53) or (54) is the required generalisation of BOLTZMANN's equation.

56. *Geometrical meaning.* Equation (53) has a simple geometrical meaning. Consider the simple particular case when $p = q = r = 0$ so that u, v, w are constants for each particle. Then (53) may be written
$$\frac{1}{(c^2 - \Sigma u^2)^{1/2}} \frac{\partial f}{\partial t} + \Sigma \frac{u}{(c^2 - \Sigma u^2)^{1/2}} \frac{\partial f}{\partial x} = 0. \tag{56}$$

But along any given world-line
$$\frac{d}{ds} = \frac{dx}{ds}\frac{\partial}{\partial x} + \frac{dy}{ds}\frac{\partial}{\partial y} + \frac{dz}{ds}\frac{\partial}{\partial z} + \frac{dt}{ds}\frac{\partial}{\partial t}$$
$$= \frac{1}{(c^2 - \Sigma u^2)^{1/2}} \left[\frac{\partial}{\partial t} + \Sigma u \frac{\partial}{\partial x}\right].$$

Hence (56) may be written
$$df/ds = 0,$$
i. e. f is constant along a world-line. Since u, v, w are here constant along the world-line, I also is constant along the world line. Thus we have shown that when the world-lines are straight lines, the intensity of world-lines is constant along a world-line, in analogy with radiation-theory. The more general equation (53) with $p \neq 0$, etc., takes account of variation in intensity of world-lines along a world-line when the path is not straight and the elementary solid angle of a pencil changes with propagation.

57. If the distribution function f given by (50) is inserted in the BOLTZMANN equation (53) we have a differential relation connecting p, q, r, ψ. There is thus a connection between the statistical distribution of position and velocity in the swarm and the accelerations of the individual members of the swarm. Put otherwise, the intensity-distribution of world-lines determines their curvature[1]). Our problem is a sort of converse of the usual radiation problem, where the refractive index is given and the intensity-distribution is required. Out of statistics we are getting dynamics. It will be shown that equation (53) can be completely integrated.

58. *Determination of the accelerations.* The accelerations p, q, r, which depend on x, y, z, t, u, v, w, are observed by all observers in their own co-ordinates. We must ensure that the observations are consistent.

Since p, q, r is a vector in 3-space, it can be resolved into components along the two 3-vectors x, y, z and u, v, w and perpendicular to them.

[1]) Thus a proposition of gravitational character is derived from purely kinematical postulates.

From considerations of symmetry, since the swarm is spherically symmetrical about $x = 0$, $y = 0$, $z = 0$, we must expect that the acceleration will be in the plane containing the origin O, the point P and the velocity. Otherwise argued, the vector p, q, r can be expressed as a linear vector function, of $\mathbf{r}\ (= x, y, z)$ and $\mathbf{v}\ (= u, v, w)$ and $\mathbf{r} \wedge \mathbf{v}$. But there is no reason why the acceleration should be on one side rather than the other of the plane of $\mathbf{r}$ and $\mathbf{v}$. Hence the coefficient of $\mathbf{r} \wedge \mathbf{v}$ must be zero[1]). Hence we write

$$p = Ax + Bv \qquad q = Ay + Bv \qquad r = Az + Bw \tag{56}$$

where A and B are 3-scalars. We wish to ensure that these are consistently transformed by all observers. For any one particle, along its track in space-time, we have

$$\frac{dx}{ds} = \frac{u}{Y^{1/2}}, \quad \frac{dy}{ds} = \frac{v}{Y^{1/2}}, \quad \frac{dz}{ds} = \frac{w}{Y^{1/2}}, \quad \frac{c\,dt}{ds} = \frac{1}{Y^{1/2}}. \tag{57}$$

We notice that

$$\frac{d\,Y^{-1/2}}{dt} = \frac{\Sigma up}{c^2\,Y^{3/2}}. \tag{58}$$

Hence

$$\frac{d^2 x}{ds^2} = \frac{dt}{ds}\frac{d}{dt}\left(\frac{dx}{ds}\right) = \frac{1}{c^2\,Y}\left[p + u\,\frac{\Sigma up}{c^2\,Y}\right], \tag{59}$$

$$c\,\frac{d^2 t}{ds^2} = c\,\frac{dt}{ds}\frac{d}{dt}\left(\frac{dt}{ds}\right) = \frac{\Sigma up}{c^3\,Y^2}. \tag{59'}$$

Now from (56)

$$\Sigma up = A\,\Sigma ux + B\,\Sigma u^2$$

which can be re-arranged in the form

$$\Sigma up = -c^2\,[AZ + BY] + c^2\,[At + B] \tag{60}$$

Substituting in equations (59) (59′) and making certain rearrangements we find

$$\frac{d^2 x}{ds^2} = \frac{A}{c^2\,Y}\left[x - \frac{u}{Y^{1/2}}\,\frac{Z}{Y^{1/2}}\right] + \frac{u}{c^2\,Y^2}\,[At + B]. \tag{61}$$

$$c\,\frac{d^2 t}{ds^2} = \frac{A}{c^2\,Y}\left[ct - \frac{c}{Y^{1/2}}\,\frac{Z}{Y^{1/2}}\right] + \frac{c}{c^2\,Y^2}\,[At + B] - \frac{1}{c\,Y}\,[At + B]. \tag{61'}$$

Now

$$\frac{d^2 x}{ds^2}, \quad \frac{d^2 y}{ds^2}, \quad \frac{d^2 z}{ds^2}, \quad \frac{c\,d^2 t}{ds^2}$$

---

[1]) This argument would not hold good if we were considering *charged* particles. Particles of opposite signs could then be accelerated to opposite sides of the plane.

must be a 4-vector[1]). But

$$x, \quad y, \quad z, \quad ct$$

and

$$u/Y^{1/2}, \quad v/Y^{1/2}, \quad w/Y^{1/2}, \quad c/Y^{1/2}$$

are 4-vectors, and $Z/Y^{1/2}$ is a 4-scalar. It follows from (61) and (61′) that A/Y must be a 4-scalar, that $(At + B)/Y^{3/2}$ must be a 4-scalar, and lastly owing to the last term in (61′) that

$$At + B = 0 \tag{62}$$

Inserting for B in (56) we have accordingly

$$p = A\,(x - ut), \qquad q = A\,(y - vt), \qquad r = A\,(z - wt) \tag{63}$$

where A/Y is a 4-scalar function of x, y, z, t, u, v, w. It is thus a function of X and Z^2/Y. A is of dimensions -2 in time and accordingly

$$\frac{A}{Y} = \frac{F\,(Z^2/X\,Y)}{X} \quad \text{or} \quad A = \frac{Y}{X}\,F\left(\frac{Z^2}{X\,Y}\right). \tag{64}$$

Formulae (63) for p, q, r will now give the same values of the curvature of the world-lines in all rectangular co-ordinate systems.

59. *Integration of the* Boltzmann *equation.* We now substitute expressions (63) in the Boltzmann equation (54) and for f we substitute its value from (50). We obtain

$$\Lambda \log\left[X^{-3/2}\,Y^{-5/2}\,\psi\left(\frac{Z^2}{X\,Y}\right)\right] + \Sigma\,(x - u\,t)\,\frac{\partial A}{\partial u} - 3\,A t = 0. \tag{65}$$

As useful lemmas we note the following:

$$\Lambda X = 2Z \tag{66}$$

$$\Lambda Y = 2A\,(Z - t\,Y) = \frac{2\,Y\,(Z - t\,Y)}{X}\,F\left(\frac{Z^2}{X\,Y}\right), \tag{66′}$$

$$\Lambda Z = Y + A\,(X - tZ) = Y + \frac{Y\,(X - t\,Z)}{X}\,F\left(\frac{Z^2}{X\,Y}\right), \tag{66″}$$

$$\Lambda \psi = 2\frac{Z}{X}\left(1 - \frac{Z^2}{X\,Y}\right)\left(1 + \frac{A\,X}{Y}\right)\psi'\left(\frac{Z^2}{X\,Y}\right) = 2\frac{Z}{X}\left(1 - \frac{Z^2}{X\,Y}\right)(1 + F)\,\psi'. \tag{66‴}$$

$$\Sigma\,(x - ut)\,\frac{\partial A}{\partial x} = -\frac{2}{X}\left[(t\,Y - Z)\,F + Z\left(\frac{Z^2}{X\,Y} - 1\right)F'\right]. \tag{66⁗}$$

---

[1]) I have seriously considered writing the whole paper in vector rotation. But as the argument alternately requires the consideration of 3-vectors and 4-vectors, it would have been necessary to devise a special notation. Many of the calculations were in fact done originally in vector notation.

When we substitute these expressions in (65) we find that the explicit terms in t all cancel. Setting

$$Z^2/XY = \xi \tag{67}$$

the equation is found to reduce to

$$\frac{3}{2}(1+F) + (\xi-1)(1+F)\frac{\psi'}{\psi} + (\xi-1)F' = 0,$$

or

$$\frac{\frac{3}{2}}{\xi-1} + \frac{\psi'}{\psi} + \frac{F'}{1+F} = 0. \tag{68}$$

The fact that ξ alone occurs in this equations (F and ψ being functions of ξ only) is a welcome check on the accuracy both of our algebra and our dimensional arguments. Equation (6) integrates as it stands in the form

$$\psi(1+F)(\xi-1)^{3/2} = \text{constant} = C$$

whence

$$F(\xi) = -1 + \frac{C}{(\xi-1)^{3/2}\psi(\xi)}. \tag{69}$$

The accelerations are accordingly

$$p = \frac{Y}{X}F(\xi)(x-ut) = -\frac{Y}{X}\left[1 - \frac{C}{(\xi-1)^{3/2}\psi(\xi)}\right](x-ut), \tag{70}$$

$$q = \frac{Y}{X}F(\xi)(y-vt) = -\frac{Y}{X}\left[1 - \frac{C}{(\xi-1)^{3/2}\psi(\xi)}\right](y-vt), \tag{70'}$$

$$r = \frac{Y}{X}F(\xi)(z-wt) = -\frac{Y}{X}\left[1 - \frac{C}{(\xi-1)^{3/2}\psi(\xi)}\right](z-wt). \tag{70''}$$

60. *The law of gravitation.* Equations (50) and (70) together constitute the law of gravitation for a system subject to the two postulates of relativity, as far as we have yet determined it. For they assert that the distribution of matter and motion given by (50) "causes" accelerations (70). The time is naturally involved, because the distribution of matter and motion is changing with the time.

We would expect that the next step would be to dissect the "global" law of gravitation (50) and (70) into "field equations" holding locally and describing the actual interactions. Actually the investigation now takes a surprising turn.

§ 10. Integration of the equations of motion. 61. *Statement of results to be proved.* We have expressed the physical conditions that the distribution is to be the same to all observers and that the distribution at one instant t shall turn into the distribution at $t + \Delta t$, *statistically*. But we have not

excluded the possibility of particles interchanging positions. The particles must actually move along the word-lines whose differential equations are

$$\frac{dx}{dt} = u, \qquad \frac{dy}{dt} = v, \qquad \frac{dz}{dt} = w, \tag{71}$$

$$\left.\begin{aligned} \frac{du}{dt} &= \frac{Y}{X}(x - ut)F(\xi), \qquad \frac{dv}{dt} = \frac{Y}{X}(y - vt)F(\xi), \\ \frac{dw}{dt} &= \frac{Y}{X}(z - wt)F(\xi). \end{aligned}\right\} \tag{71'}$$

Integration of equations (71) and (71′) should give x, y, z, u, v, w as functions of t involving 6 arbitrary constants; these should be the equations of the world-lines. It may seem a surprising thing that equations (71) and (71′) can be integrated without knowing the explicit form of $F(\xi)$. We shall now however prove that the only integrals of (71 (71′) are

$$u = \text{constant}, \qquad v = \text{constant}, \qquad w = \text{constant}, \tag{72}$$

with x, y, z linear functions of t, and that accordingly

$$F \equiv 0. \tag{73}$$

This determines the distribution-law in detail, for it gives from (70)

$$\psi(\xi) = \frac{C}{(\xi - 1)^{3/2}} = \frac{C\, X^{3/2}\, Y^{3/2}}{(Z^2 - XY)^{3/2}}, \tag{74}$$

whence from (50)

$$f(x, y, z, t, u, v, w)\, dx\, dy\, dz\, du\, dv\, dw = \frac{C\, dx\, dy\, dz\, du\, dv\, dw}{c^6\, Y\, (Z^2 - XY)^{3/2}}. \tag{75}$$

Since identically

$$Z^2 - XY \equiv \frac{1}{c^2}\left[\Sigma (x - ut)^2 - \frac{1}{c^2}\Sigma (vz - wy)^2\right], \tag{76}$$

$$\equiv \frac{1}{c^2}\left[\Sigma (x - ut)^2 - \frac{1}{c^2}\Sigma \{v(z - wt) - w(y - vt)\}^2\right], \tag{76'}$$

we see by inspection that distribution (75) satisfies the physical condition implied by Boltzmann's equation; for if $u = u_0$, $v = v_0$, $w = w_0$, the value of f is unaltered by increasing x by $u_0 T$, y by $v_0 T$, z by $w_0 T$ when t increases by T, by (76′). Hence a particle at (x, y, z) at time t is found at $x + u_0 T$, $y + v_0 T$, $z + w_0 T$ at time $t + T$.

62. The integration is effected by considering the terms of the mixed tensor $T^{\alpha}{}_{\beta}$,

$$T^{\alpha}{}_{\beta} = \overset{\alpha}{\downarrow} \overset{\beta \rightarrow}{} \begin{matrix} 0, & \dfrac{uy - vx}{Y^{1/2}}, & \dfrac{uz - wx}{Y^{1/2}}, & \dfrac{c(x - ut)}{Y^{1/2}}, \\ \dfrac{vx - uy}{Y^{1/2}}, & 0, & \dfrac{vz - wy}{Y^{1/2}}, & \dfrac{c(y - vt)}{Y^{1/2}}, \\ \dfrac{wx - uz}{Y^{1/2}}, & \dfrac{wy - vz}{Y^{1/2}}, & 0, & \dfrac{c(z - wt)}{Y^{1/2}}, \\ \dfrac{c(x - ut)}{Y^{1/2}}, & \dfrac{c(y - vt)}{Y^{1/2}}, & \dfrac{c(z - wt)}{Y^{1/2}}, & 0. \end{matrix}$$

It is readily verified that $T^{\alpha}{}_{\beta}$ is connected with the tensor

$$g^{\alpha\mu} E_{\beta\mu\gamma\delta} r^{\gamma} v^{\delta}$$

with $g_{\alpha\mu}$ given by our ds^2, $r^{\gamma} = (x, y, z, ct)$ and $v^{\delta} = (u/Y^{1/2}, v/Y^{1/2}, w/Y^{1/2}, c/Y^{1/2})$, E being the corresponding "alternate" tensor. Write

$$f = (x - ut)/Y^{1/2}, \qquad g = (y - vt)/Y^{1/2}, \qquad h = (z - wt)/Y^{1/2}, \tag{77}$$

$$l = (vz - wy)/Y^{1/2}, \quad m = (wx - uz)/Y^{1/2}, \quad n = (uy - vx)/Y^{1/2}. \tag{77'}$$

63. Then by actual differentiation we find using (71) and (71')

$$\frac{df}{dt} = -f \frac{Z}{X} F(\xi) \tag{78}$$

the terms involving t explicitly cancelling identically. Similarly

$$\frac{dl}{dt} = -l \frac{Z}{X} F(\xi). \tag{78'}$$

We have accordingly

$$\frac{d(\log f)}{dt} = \frac{d(\log g)}{dt} = \frac{d(\log h)}{dt} = \frac{d(\log l)}{dt}$$
$$= \frac{d(\log m)}{dt} = \frac{d(\log n)}{dt} = -\frac{Z}{X} F(\xi). \tag{79}$$

This gives five integrals in the form

$$f : g : h : l : m : n = \text{constants}. \tag{80}$$

It is convenient to equate each of expressions (79) to

$$\frac{d \log K}{dt},$$

so that

$$\frac{d \log K}{dt} = -\frac{Z}{X} F(\xi). \tag{81}$$

A sixth integral can now be obtained. By direct differentiation along a world-line we have

$$\frac{dX}{dt} = 2Z, \tag{82}$$

$$\frac{d\xi}{dt} = 2\frac{Z}{X}(1-\xi)[1+F(\xi)]. \tag{83}$$

Hence

$$\frac{d\log K}{d\xi} = \frac{1}{2}\frac{F(\xi)}{(\xi-1)[1+F(\xi)]}, \tag{84}$$

$$\frac{1}{X}\frac{dX}{d\xi} = -\frac{1}{(\xi-1)[1+F(\xi)]}, \tag{85}$$

whence

$$\frac{d\log K}{d\xi} - \frac{1}{2}\frac{d\log X}{d\xi} = \frac{\frac{1}{2}}{\xi-1},$$

which integrates in the form

$$\left.\begin{aligned} K &= D(\xi-1)^{1/2}X^{1/2}, \\ &= D\left(\frac{Z^2}{Y}-X\right)^{1/2}, \end{aligned}\right\} \tag{86}$$

whilst (85) gives

$$X = E\exp.\left[-\int^{\xi}\frac{d\xi}{(\xi-1)[1+F(\xi)]}\right]. \tag{87}$$

In (86) and (87) D and E are constants of integtation. We can now write integrals (80) in the form

$$\left.\begin{aligned} x-ut &= f_0(Z^2-XY)^{1/2}, \\ y-vt &= g_0(Z^2-XY)^{1/2}, \\ z-wt &= h_0(Z^2-XY)^{1/2}, \end{aligned}\right\} (88) \qquad \left.\begin{aligned} vz-wy &= l_0(Z^2-XY)^{1/2}, \\ wx-uz &= m_0(Z^2-XY)^{1/2}, \\ uy-vx &= n_0(Z^2-XY)^{1/2}, \end{aligned}\right\} (89)$$

where $f_0, g_0, h_0, l_0, m_0, n_0$ are constants of integration.

We have apparently obtained 7 integrals, namely (87), (88) and (89) so there must be at least one relation between them. In virtue of identity (76) we have the relation

$$c^2 = \Sigma f_0^2 - \frac{1}{c^2}\Sigma l_0^2 \tag{90}$$

and further since $\Sigma(x-ut)(vz-wy) \equiv 0$ we have

$$\Sigma l_0 f_0 = 0. \tag{91}$$

As we have two relations between the 6 constants of integration $l_0, \ldots h_0$, equations (88) and (89) are equivalent to only 4 independent integrals. Taking into account (87) we therefore expect still one further constant to appear.

Substituting for x, y, z from (88) in (89) we have

$$\left.\begin{aligned} h_0 v - g_0 w &= l_0, \\ f_0 w - h_0 u &= m_0, \\ g_0 u - f_0 v &= n_0 \end{aligned}\right\} \tag{92}$$

the factor $(Z^2 - XY)^{1/2}$ dividing out. [If $Z^2 - XY = 0$ we have the integrals which are the world lines $u = u_0$, $v = v_0$, $w = w_0$, $x = u_0 t$, $y = v_0 t$, $z = w_0 t$, a series of radial straight lines traversed with uniform velocity[1]).] The solution of (92) is

$$\left.\begin{aligned} u &= \frac{g_0 n_0 - h_0 m_0}{f_0^2 + g_0^2 + h_0^2} + f_0 \lambda(t), \\ v &= \frac{h_0 l_0 - f_0 n_0}{f_0^2 + g_0^2 + h_0^2} + g_0 \lambda(t), \\ w &= \frac{f_0 m_0 - g_0 l_0}{f_0^2 + g_0^2 + h_0^2} + h_0 \lambda(t), \end{aligned}\right\} \tag{93}$$

where $\lambda(t)$ is an undetermined function of t. Hence by (88)

$$x = \frac{g_0 n_0 - h_0 m_0}{f_0^2 + g_0^2 + h_0^2} t + f_0 t \lambda(t) + f_0 (Z^2 - XY)^{1/2} \tag{94}$$

with two similar equations for y and z. Since (93) and (94) must satisfy (71) we must have

$$\frac{d}{dt}(Z^2 - XY)^{1/2} = -t\lambda'(t),$$

whence

$$(Z^2 - XY)^{1/2} = -\int_0^t t\lambda'(t) + a \tag{95}$$

a being constant along the world-line in question.

Now $(Z^2 - XY)^{1/2}$ or $Y^{1/2} X^{1/2} [(Z^2/XY) - 1]^{1/2}$ is clearly covariant with $Y^{1/2}$ under a Lorentz transformation. Hence the right-hand side of (95) must be independent of t, and so $\lambda'(t) = 0$, $\lambda(t) = b$; a and b with $f_0, \ldots n_0$ form 6 *independent* constants of integration.

The integrals (93) and (94) may now be written

$$\left.\begin{aligned} u &= u_0, & v &= v_0, & w &= w_0, \\ x &= u_0 t + x_0, & y &= v_0 t + y_0, & z &= w_0 t + z_0. \end{aligned}\right\} \tag{96}$$

By evaluating $x - ut$, $vz - wy$ and $Z^2 - XY$ we find

$$(Z^2 - XY)^{1/2} = \frac{1}{c}\left[\Sigma x_0^2 - \frac{1}{c^2}\Sigma (v_0 z_0 - w_0 y_0)^2\right],$$

$$f_0 = \frac{x_0}{(Z^2 XY)^{1/2}}, \quad l_0 = \frac{v_0 z_0 - w_0 y_0}{(Z^2 - XY)^{1/2}},$$

---

1) In this exceptional case, which proves in § 13 to be of fundamental importance, we cannot infer $F \equiv 0$.

so that (90) and (91) are satisfied identically. Further the integral relation (87) is now satisfied identically. For since u, v, w are constants, $F(\xi) \equiv 0$, and (87) gives

$$X = \frac{E}{\xi - 1} = \frac{EXY}{Z^2 - XY},$$

which is true since $Z^2 - XY$ and Y are constants. This is the desired result.

64. *Implications of the postulates.* We have now obtained, without any dynamical postulates a dynamical result, namely that the paths of all particles of the system subject to the two postulates of relativity, are straight lines described with uniform velocities. We have also obtained, in the formula (75), the complete distribution of particle-density and particle velocity in the system.

It will be seen that what we have established can be stated as follows:

If it is postulated: (1) that the velocity of light is constant as observed by all observers in uniform motion with respect to one another, (2) that the distribution of the particles of the system in position and in velocity shall exhibit identical sequences of configurations to all observers "similarly oriented" with respect to the system, whereever they be, then every particle of the system must describe a straight line with uniform velocity and the distribution of particles in position is given by (75). By "similarly oriented" it is meant that each observer must assume the velocity which makes the system appear to be centred round himself, this being always possible. The origin of x, y, z is then the position of the observer, and the origin of t is that epoch at which any two similarly oriented observers must have parted company.

Part of this result can be stated in the form that the two postulates imply: (1) Newton's first law of motion for all particles in the world; (2) absence of gravitation.

Since gravitation is a constituent of the real world, we infer that both postulates cannot hold good in the real world. As we are philosophically unwilling to surrender the second postulate — for this would be to attach a meaning to absolute position in the world — we must surrender the first. Hence the velocity of light cannot be constant, and so gravitation must have an effect on light.

We can however establish that gravitation *must* be a constituent of the real world, for the spatio-velocity distribution (75) involves singularities and we cannot have singularities in the real world. We now examine the nature of the spatio-velocity distribution we have determined.

§ *11. Nature of the spatio-velocity distribution.* 65. *Properties at* $t = 0$. At the instant $t = 0$ in any observer's reckoning, the distribution-law (75) gives

$$\frac{C\,dx\,dy\,dz\,du\,dv\,dw}{c^6\left(1-\frac{\Sigma u^2}{c^2}\right)\left[\left(\frac{\Sigma u\,x}{c^2}\right)^2+\left(\frac{\Sigma x^2}{c^2}\right)\left(1-\frac{\Sigma u^2}{c^2}\right)\right]^{3/2}}.$$

Consider the particles moving inside the shell $(r, r+dr)$ in a given direction with a given velocity. Let V be the given velocity (not to be confused with our previous use of V for an observer's velocity); then $du\,dv\,dw = V^2 dV\,d\omega$. Choose the axis of x to be along the direction of the given velocity. Put

$$x = r\cos\vartheta, \qquad y = r\sin\vartheta\cos\varphi, \qquad z = r\sin\vartheta\sin\varphi. \tag{97}$$

In this (x, y, z) frame, the velocities u, v, w are given by

$$u = V, \qquad v = 0, \qquad w = 0.$$

Hence the number of particles inside $(r, r+dr)$ moving with V inside $dV\,d\omega$ is

$$\frac{C\,V^2\,d\omega\,dV}{c^6\,(1-V^2/c^2)}\,\frac{r^2\,dr}{r^3}\int\limits_{\varphi=0}^{2\pi}\int\limits_{\vartheta=0}^{\pi}\frac{\sin\vartheta\,d\vartheta\,d\varphi}{\left[\frac{V^2\cos^2\vartheta}{c^4}+\frac{1}{c^2}\left(1-\frac{V^2}{c^2}\right)\right]^{3/2}}$$

which works out to be

$$\frac{C\,V^2\,dV\,d\omega}{c^3\,(1-V^2/c^2)^2}\cdot\frac{4\pi r^2\,dr}{r^3}. \tag{98}$$

This shows that at $t = 0$, when the velocity-distribution is considered as summed over a spherical shell it is independent of the position of the spherical shell, being of the type

$$\frac{V^2\,dV\,d\omega}{(1-V^2/c^2)^2}. \tag{99}$$

This is precisely of the type found in § 7 and gives a range of velocities up to c. The spatial distribution is by (98) of the type

$$\varrho \propto 1/r^3. \tag{100}$$

This gives an infinity in the density at $r = 0$, and a decreasing density outside.

66. *Convergence of the gravitational potential at infinity.* The fact that at $t = 0$ the density decreases as the inverse cube of the radius is of funda-

mental importance. For this characteristic will be expected to be only slightly modified by gravitation, if a similar analysis were attempted on more general lines, i. e. by allowing for the effect of gravitation on light and so replacing the Lorentz transformation by something more general. But the Newtonian law

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dV}{dr}\right) = -4\pi G\varrho$$

with $\varrho \propto 1/r^3$ gives at once a potential of the form

$$V \propto A\frac{\log r}{r} + \frac{B}{r^2} + \frac{C}{r} + D \qquad (101)$$

which converges as $r \to \infty$.

Thus the distribution $\varrho \propto 1/r^3$ gives a gravitational potential converging at infinity, and so avoids the difficulty concerning a *uniform* density-distribution brought forward by EINSTEIN.

67. *Expansion phenomena.* Since at time $t = 0$ the velocity-distribution involves only V^2, it follows that both the velocity-distribution itself and its velocity-reverse give rise to expansion. Hence the instant $t = 0$ is also the „natural origin" of time for this system, in the kinematic sense of § 3.

The motion from $t = 0$ onwards is accordingly one of natural expansion. The distribution at any later time may be easily inferred without detailed calculation. We have first the natural motion of the cloud of particles with spatio-velocity distribution (98). [The gravitational potential at infinity (for a similar swarm subject to gravitation) may be expected to continue to converge; for it will alter continuously with the time.] We have secondly the spatio-velocity distribution brought about by the presence of the singularity at $r = 0$ at time $t = 0$. At any time $t > 0$ particles which at time $t = 0$ were at the origin and moving with velocity V will have arrived at a distance $r = Vt$, and the sphere $r = Vt$ will thus be a locus of singularities for particles moving radially with speed V. This is easily seen from (75) when identity (76') is taken into account. For the distribution function has a singularity at $x = ut$, $y = vt$, $z = wt$. In other words, at any given point in space, say P at distance r from 0, at any assigned moment t there exists a speed V equal to r/t such that particles moving radially outwards with the speed V at P at the instant t have an infinity in their partial density. Thus the initial singularity at $r = 0$ at time $t = 0$ is a source of singularities, which move outwards at all speeds $V < c$. The origin $r = 0$ is always a singularity, namely for particles at rest there.

Thus if we consider particles moving with an assigned *velocity* V, given in direction and magnitude, the density of such particles has a singularity i. e. a maximum, at the point $r = Vt$. Everywhere else there is no singularity for such particles.

There are no singularities whatever outside the sphere $r = ct$ at any instant $t > 0$.

Any other observer anywhere else, moving with the velocity which would just have brought him from the origin in time t, sees exactly the same spatio-velocity distribution. For example the singularity at $r = 0$ for resting particles in the frame of the observer A at $r = 0$ is regarded as an "$r = Vt$" singularity by an observer B at $r = Vt$. What was regarded as an eccentric singularity by A becomes the origin-singularity to B. Each observer sees singularities of this type extending up to $r = ct$, outside $r = ct$ no singularities. That each observer regards himself an the centre of the sphere $r = ct$ is of course a fundamental result in the existing special relativity theory. Our spatio-velocity distribution is a generalisation of this.

The number of particles moving in a given direction with velocity V, at time t, inside the spherical shell $(r, r + dr)$ is clearly

$$\frac{C\, V^2\, dV\, d\omega}{c^6\,(1 - V^2/c^2)} \cdot r^2\, dr \cdot \int\limits_{\varphi=0}^{2\pi} \int\limits_{\vartheta=0}^{\pi} \frac{\sin\vartheta\, d\vartheta\, d\varphi}{\left[\left(t - \frac{V r \cos\vartheta}{c^2}\right)^2 - \left(t^2 - \frac{r^2}{c^2}\right)\left(1 - \frac{V^2}{c^2}\right)\right]^{3/2}}.$$

This integral can be evaluated. Two cases arise, according as $r < Vt$ or $r > Vt$. The value is found to be

$$(1)\ (r < V t);\ \frac{C\, V^2\, dV\, d\omega}{c^3\,(1 - V^2/c^2)} \cdot \frac{t\, 4\pi r^2\, dr}{V\,(t^2 - r^2/c^2)\,(V^2 t^2 - r^2)} \tag{102}$$

$$(11)\ (r > V t);\ \frac{C\, V^2\, dV\, d\omega}{c^3\,(1 - V^2/c^2)^2} \cdot \frac{4\pi r^2\, dr}{r\,(r^2 - V^2 t^2)}. \tag{103}$$

Formula (103) reduces to (98) for $t = 0$. These formulae naturally possess singularities at $r = Vt$.

We may say that owing to these singularities, at any point P *inside* $r = ct$ the velocity is predominantly one of radial expansion from O with velocity OP/t. Outside $r = ct$ the velocity-distribution is still an expanding one, but less predominantly. Particles at any given point P moving other than radially outward with speed OP/t are numerically inconsiderable compared with the directly receding particles, inside $r = ct$.

We have not assumed that $t = 0$ was the moment of creation of the system, and indeed the particles constituting the cloud $\varrho \propto 1/r^3$, at time $t = 0$, outside the point $r = 0$ itself, can all be traced back to negative times. But all the particles constituting the singularities may be said to have been created at the origin at the instant $t = 0$, for they have all originated from that point at that instant.

§ 12. General effect of gravitation. 68. *Removal of singularities.* I see no reason in principle why the investigation I have now completed within the framework of special relativity should not be capable of generalisation to include the effects of gravitation[1]). We should seek a spatio-velocity distribution in which all the light-paths are no longer straight lines — a distribution in which light is subject to gravitation in a way which it would be one of the objects of the problem to determine. The Lorentz transformation would then have to be generalised so as to give consistent pictures of the universe to any two observers, in suitable relative motion, who communicate with one another and observe distant objects by means of light-signals, the transmission process itself being subject to gravitation. With this new group of transformations the analysis of the present paper would then have to be repeated. The postulate of special relativity would be replaced by the postulate that the velocity of light as observed by the observer *near himself* must be the same *for all observers*[2]). The extended postulate would be retained. The additional condition would ultimately be imposed that the final spatio-velocity distribution be free from singularities.

Whether this programme can be carried out must be left to the future. In the meantime it is possible to see in a general way how gravitation may be expected to modify the spatio-velocity distribution here obtained.

69. The general effect will be to remove singularities, by smoothing them out. Singularities may be expected to be replaced by density-maxima.

We shall therefore expect to find, in the frame of any one observer, a maximum density at the origin and a continuum of maxima up to $r = ct$, the maximum at, say, a point P comprising those particles which are endowed with a velocity OP/t and move radially outwards. Their velocities will range up to c.

---

[1]) § 13 written later, takes complete account of the effects of gravitation. But as it is only readily intelligible in the light of § 12, I leave § 12 as originally written.

[2]) Though possibly changing secularly with the time as the universe expanded.

Superimposed on this characteristic there will be a "clotting" effect. Certain particles will be moving with such large velocities relative to an observer moving with a given velocity that they will escape from him; The remainder will not escape. The result will be, as explained in § 3, to give rise to a set of closed sub-systems all escaping from one another with velocities ranging up to c.

The two effects together suggest that the sub-systems will have a strong preference for directly radial outward motion, and of such a kind that at a point P, at the instant t in the time-scale of an observer at rest at the origin, the velocity of recession will be OP/t. The points of clotting, the centres of mass of the sub-systems, may be considered the "lattice points" — the "Gitterpunkte" — of space. The question arises as to the sizes of the sub-systems and their distribution in space and velocity. The following considerations suggest the probable answers to these problems[1]).

70. When the sub-systems have become formed out of the original distribution of particles, and have had time to separate from one another — which they must necessarily do granted sufficient time — they will constitute a system of which the components move with approximately uniform velocities without collisions. They will therefore be akin to the system of particles of § 7. Retaining the considerations arising from the extended principle of relativity, we shall then expect for this system of quasi-independent *sub-systems* (in spite of the presence of gravitation) the velocity-distribution found in § 7 for a system of collisionless rectilinearly moving particles, namely

$$\frac{A'\,du\,dv\,dw}{c^3\,(1 - \Sigma\,u^2/c^2)^2}. \tag{104}$$

In other words, the system of gravitationally closed sub-systems may be expected to mimic, in its velocity-distribution, the primeval system of particles out of which it gravitationally evolved. The constant A' will be much smaller than the original constant A, the ratio A/A' being the number of the particles comprised in a sub-system. This ratio A'/A will depend essentially on the nature of gravitation — on the absolute magnitude of the gravitational accelerations existing between the particles. We can however expect that each sub-system will contain the same number of particles. For there is no reason why one system should be bigger than any other; we can choose *any one* system as the arbitrary one to which the observer is attached; i. e. any one system as the first one (constructed out of the original assembly) of the infinite series of sub-systems.

[1]) The rigorous theory is given in § 13.

Now the spatial distribution of the sub-systems will necessarily be very different from the original distribution of particles. We have seen reason to expect that the sub-systems will be formed in the positions of the *singularities* in the primeval distribution. Since these *singularities* move outwards according to the law $r = Vt$ we shall expect that the velocity-distance proportionality will not only be obeyed *statistically* (as it was for the primeval system) but also *strictly*, for each individual sub-system. Hence the velocity distribution law is equivalent to a space-distribution law. The law

$$\frac{A' V^2 \, dV \, d\omega}{c^3 (1 - V^2/c^2)^2} \tag{104'}$$

with $r = Vt$ is in fact equivalent to the space distribution law

$$\frac{A' r^2 \, dr \, d\omega}{c^3 t^3 (1 - r^2/c^2 t^2)^2}, \quad \text{or} \quad \frac{A' t \, r^2 \, dr \, d\omega}{c^3 (t^2 - r^2/c^2)^2} \tag{105}$$

which is the number of sub-systems inside the solid angle $d\omega$ inside the shell $(r, r + dr)$ at time t.

Formula (105) gives an infinite number of sub-systems in total, the sphere $r = ct$ being a sphere of condensation — a locus of limiting points of the "open" set formed by the sub-systems.

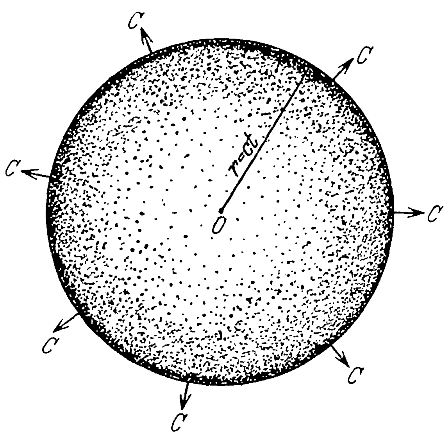

Fig. 3.

71. But it must be particularly noted that the sub-systems moving with velocity approaching c are essentially unobservable — invisible. For they are receding so fast that their apparent brightness is reduced very considerably, the factor being in fact $(1 - V/c)^2$, which for $V \sim c$ is zero. By a somewhat long calculation[1]), I have determined the amount of light (in energy units) which between epochs t and $t + dt$ (in the time-frame of one observer at the centre O) would fall on a photographic plate of area dS at O, having originated from a receding nebula in direction i, as

$$\frac{\left(\int I'_{\nu'} \, d\nu'\right) dS_0 \, dt \, (i \cdot dS) \, (1 - V/c)^2}{V^2 t^2} \tag{106}$$

where dS_0 is the projected area of the nebula normal to the direction i, and $I'_{\nu'} d\nu'$ is its specific intensity of radiation in frequency ν' measured

[1]) See appendix.

in the frame of an observer moving with it. $\int I'_{\nu'} d\nu'$ is thus a physical constant characteristic of a sub-system, the same (on the extended principle of relativity) for all sub-systems[1]). It follows that the closer the velocity V is to c, the fainter the nebula as recorded on a plate fixed at the centre of the system. In the limit, at distance ct, the nebulae are totally invisible.

We have accordingly for the total light received from sub-systems lying inside $d\omega$ the product of the last two formulae given. We see that the light received from inside $d\omega$ from sub-systems moving with recession velocity V is proportional to

$$\left.\begin{aligned} &\frac{V^2\, dV\, d\omega}{(1 - V^2/c^2)^2} \cdot \frac{(1 - V/c)^2}{V^2 t^2} \\ \text{i. e. to} \quad &\frac{dV\, d\omega}{t^2 (1 + V/c)^2} \end{aligned}\right\} \qquad (107)$$

The integral of this up the limit $V = c$ therefore converges, and the total light received is finite, in accordance with the observation that the brightness of the night sky is finite.

Outside the sphere $r = ct$ there will be the same process of "clotting" at work — the same tendency to form sub-systems; but since no singularities occur in this region in the case of gravitation absent, we shall expect the resulting sub-systems to contain very small numbers of particles, and be of an altogether smaller order of magnitude than the sub-system inside $r = ct$. The density will fall off with distance according to law (103) — a law which express the natural expansion of an originally inverse-cube distribution. But the effect of the infinite mass of the infinite number of sub-systems inside $r = ct$ will be, eventually, to sweep the outer region empty.

The resulting picture of the world is highly interesting. The overwhelming majority of the spiral nebulae — if we identify our sub-systems with spiral nebulae — may be expected to be concentrated towards the inner surface of an expanding sphere of radius $r = ct$ centred at the observer. This system forms in mathematical terminology an "open" set of points with every point of the sphere $r = ct$ a limiting point. The density (no. of nebulae per unit volume) increases up to this limiting sphere, but here the nebulae are moving with velocities approaching light and suffer a considerable Lorentz contraction in the radial direction. They suffer also a diminution in apparent brightness owing to their radial motion. It follows that if the

---

1) *Note added in proof.* This requires modification. $\int I'_{\nu'}\, d\nu'$ will be a function of t' — the same function for all nebulae — where t' is the time kept by an observer moving with the nebula. As the nebula evolves the integral may change.

distance of a nebula is judged by its brightness and its brightness alone, too large a distance will be assigned to it. Each nebula must be brought forward by an amount which depends on the nearness of its velocity to c. The very faint, very fast moving nebula must be brought nearer by a very appreciable factor as compared with their uncorrected distances. I have had no opportunity of examining the Mount Wilson and other plates containing it is said countless numbers of faint nebulae but our picture of the universe is not in contradiction with this observation. It has been held that the distribution of density is a uniform one. But this has been derived in the belief that the distance of a nebula could be directly inferred from its brightness. When account is taken of the reduction in observable brightness caused by recession, the distribution previously considered spatially uniform must be compressed, the more distant ones being brought nearer. A spatially uniform distribution may be analysed as

$$\varrho r^2\, dr\, d\omega \text{ nebulae each of } \textit{apparent} \text{ brightness } A/r^2,$$

i. e. using the velocity-distance proportionality,

$$\varrho V^2\, dV\, d\omega \text{ nebulae each of apparent brightness } A/V^2.$$

This is the old interpretation. My interpretation is that there are:

$$\frac{V^2\, dV\, d\omega}{(1 - V^2/c^2)^2} \text{ nebulae each of apparent brightness } \frac{A\,(1 - V/c)^2}{V^2}.$$

In each case the number of nebulae *observed* in $dV\, d\omega$ is inversely as the observed brightness of an individual system[1]). Thus the old observation previously interpreted as indicating a *uniform* distribution of nebulae is fully compatible with my interpretation as a condensation of nebulae towards the inside of the sphere $r = ct$. I consider this a fundamentally important point in which my theory is confirmed by observation.

72. The accelerations of the particles will no longer be strictly zero[2]). I propose to examine in a general way what will be the *value* of the acceleration in the vicinity of the *centre* of the whole system of nebulae, i. e. near the observer. Since the accelerations of the nebulae are no longer strictly zero, the function F will no longer be rigorously zero[3]). Consider then the form which our acceleration formulae (70) will probably take for *small* values of x, y, z (i. e. near the observer) for a particle moving with *small* velocity u, v, w. We have then $X \sim t^2$, $Y \sim 1$, $Z \sim t$, $\xi = Z^2/XY$

---

[1]) Apart from an unimportant factor $(1 + V/c)^2$.

[2]) But in the exact solution of § 13 the accelerations are shown to be strictly zero. The relations between local gravitation and gravitation on the large scale must be left to another paper for fuller discussion.

[3]) One former result $F \equiv 0$ is invalid precisely in the case $r = Vt$, as we saw.

~ 1. The factor $x - ut$ may be expected to undergo only a very small change of form. Necessarily we must now take u small compared with x/t, so that $x - ut$ is approximately x. We have then

$$p \sim -\frac{x}{t^2}[-F(1)], \quad q \sim -\frac{y}{t^2}[-F(1)], \quad r \sim -\frac{z}{t^2}[-F(1)] \tag{108}$$

where $-F(1)$ is a small constant.

But this vector p, q, r should represent the gravitational acceleration near the centre of the world, i. e. at a small distance (x, y, z) from the observer, reckoned in his frame. Now we ordinarily describe the gravitational acceleration a small distance R from a central observer as approximately

$$G\frac{\frac{4}{3}\pi\varrho_0 R^3}{R^2} \tag{109}$$

where G is the Newtonian constant of gravitation and ϱ_0 the mean density of the (smoothed-out) universe near the observer. The acceleration (109) has components

$$-\tfrac{4}{3}\pi G\varrho_0 x, \quad -\tfrac{4}{3}\pi G\varrho_0 y, \quad -\tfrac{4}{3}\pi G\varrho_0 z. \tag{110}$$

Identifying (110) with the components p, q, r of (108) we have

$$\frac{4}{3}\pi G\varrho_0 = \frac{-F(1)}{t^2}. \tag{111}$$

Since $-F(1)$ may be expected to be not large, (111) gives approximately

$$\frac{4}{3}\pi G\varrho_0 < \frac{1}{t^2} \tag{112}$$

which is accordingly an inequality to be satisfied between G the "constant" of gravitation, ϱ_0 the mean density of the universe (smoothed out) near the observer and $1/t$, the coefficient of proportionality in the velocity-distance law. Taking $G = 6{,}66 \cdot 10^{-8}$ and using the value of t found for the real universe in § 3, para 12, we find from this

$$\varrho_0 < \frac{1}{\frac{4}{3}\pi \cdot 6{,}66 \cdot 10^{-8} \cdot (0{,}6 \cdot 10^{17})^2} = 10^{-27}\ \text{g.} \cdot \text{cm.}^{-3} \tag{113}$$

73. Relation (111) or inequality (112) replaces on the kinematic theory the relations between the mean-density of the universe, the "radius of curvature of the universe" and the coefficient of proportionality in the velocity-distance law given by the "expanding space" theory. Substantially (111) is the same as that given by the "expanding-space" theory if we eliminate from the relations occuring in that theory the unobservable and meaningless "radius of curvature" of the universe. The latter has been seen to be an unnecessary element in the situation, logically, and we now

find that we can perform our calculations just as well without it. Moreover our result has the satisfactory feature that it merely affirms that the mean density of the smoothed-out universe *near the observer should not exceed* 10^{-27} gram · cm.$^{-3}$.

Formula (111), obtained solely by kinematical arguments, has a simple dynamical interpretation. According to our ordinary ideas we should say of two neighbouring mutually receding nebulae that the one possesses a velocity exceeding the velocity of escape from the "field" of the other. If v is the relative velocity, a the separation, M the mass, this means

$$\frac{1}{2} v^2 > \frac{GM}{a}.$$

But $v = \frac{a}{t}$. Hence

$$\frac{1}{2} \cdot \frac{1}{t^2} > \frac{GM}{a^3} \sim G \varrho_0, \tag{112'}$$

Thus the relation between the mean-density ϱ_0 the smoothed-out universe near the observer and the coefficient $1/t$ in the recession law, obtained in so indirect a way on the "expanding-space" theory *via* the metric, is nothing more or less than the statement that neighbouring nebulae are escaping from one another's Newtonian fields.

74. It may be remarked at this stage that if the world actually began at time $t = 0$, the universe *accessible to observation* for a properly oriented observer will possess an actual radius of curvature, namely the length ct. But this is not a radius of curvature of "space" — the space is flat — it is the radius of the spherical volume accessible to observation at the present epoch. This of course increases with time t at the rate c. Its value is simply as many light years as t is years, namely $2 \cdot 10^9$. This is of course of the order of the value assigned on the "expanding space" theory to the mysterious "radius of curvature of space". The distribution of matter outside $r = ct$ if any exist, is essentially unobservable.

§ 13. Hydrodynamic treatment[1]). 75. In § 12 I have sketched the probable effects to be expected from the influence of gravitation on a moving swarm whose statistics satisfied the two postulates of relativity. We were led to infer, purely from physical considerations, that after gravitation had caused the formation of clots or sub-systems, the velocity-statistics for the sub-systems would be the same as those for a system of particles, namely (104)

[1]) I may perhaps be allowed to state that the investigations of § 13 were carried out only after all the preceding work had done and written up as it now stands.

and since we expected the clots or sub-systems to be formed in the spatial positions of the singularities in the original swarm, we were led to infer an *exact* correlation between position and velocity according to the law $V = r/t$; this being contrasted with the merely *statistical* correlation shown in *PtI* to hold for *any* system whose properties were chiefly determined by its kinematics. Formula (104) for the velocity-distribution together with $V = r/t$ gave formula (105) for the space distribution and its alteration with time.

In these days when physical considerations are scouted as of no importance and mathematical symbolism is held to be both our only resource and our only objective, I think it of some interest to let the above suggestions and ideas stand in the form in which they first occurred to me. For they give a simple, clear, unambiguous picture of the universe easily expounded without any appeal to geometries of space or to a supposed world geometry. But the resulting world-picture can be derived by direct mathematics in the following way. The present section, § 13, which contains the principal result of the paper as described in the summary, is logically independent of the rest of the paper.

76. It is clear that once we introduce an *exact* correlation between velocity and position, our system is not a statistical one but a hydrodynamical one. Instead of there being a distribution of velocity at each point, there is a unique velocity at each point. The motion is now one of *flow*. The usual conceptions of stream-lines and tubes of flow are therefore available. It accordingly suggests itself to ask the following question: what is the most general possible hydrodynamical motion of flow consistent with the two postulates of relativity? Or, what is the most general possible *systematic* motion in the world compatible with the constancy of the velocity of light and the postulate that the motion of flow must appear the same to all observers in uniform motion with respect to one another? I propose now to investigate this problem. We shall find that the answer is that the motion of flow is precisely that given by (104) and (105).

77. A hydrodynamical motion of flow is defined by a density ϱ[1]) and a velocity u, v, w, each of the four variables ϱ, u, v, w being functions of position and time which are subject to the equation of continuity

$$\frac{\partial \varrho}{\partial t} + \frac{\partial}{\partial x}(\varrho u) + \frac{\partial}{\partial y}(\varrho v) + \frac{\partial}{\partial z}(\varrho w) = 0. \tag{114}$$

[1]) By density ϱ I mean of course the number of particles per unit volume. No idea of mass is involved. It would have been better to describe the system as *hydrokinetic*.

The latter is the only condition of consistency required. It holds in the frame of an arbitrary observer.

Now let the four functions be

$$\varrho = \chi\,(x, y, z, t),\ u = \alpha\,(x, y, z, t),\ v = \beta\,(x, y, z, t),\ w = \gamma\,(x, y, z, t) \quad (115)$$

as measured by an observer A in his own frame. At the instant $t = 0$ in A's frame let a second observer B set off from A with uniform velocity V along A's x-axis, and let him determine also the flow in his (B's) frame. Let the result of his observations be

$$\begin{gathered}\varrho = \chi_1\,(x, y, z, t),\\ u = \alpha_1\,(x, y, z, t),\ v = \beta_1\,(x, y, z, t),\ w = \gamma_1\,(x, y, z, t)\end{gathered} \quad (115')$$

as measured by B in his own frame.

By the first postulate of relativity (constancy of the velocity of light) the Lorentz formulae connect the observations of an event in A's and B's frames. We must obtain the conditions that (115) and (115′) give consistent descriptions of the world.

Plot the paths of the particles in a *map* in which the metric is the invariant $c^2\,dt^2 - dx^2 - dy^2 - dz^2$. Then the world lines will build up a system of world-tubes, since there is a determinate direction (u, v, w) at each point (x, y, z, t) in the map in any observer's frame. Consider such a world-tube in the vicinity of a point (x, y, z, t) and let $d\Sigma$ be its normal cross-section here. Then the intensity J of world lines at (x, y, z, t) may be defined as such that $J\,d\Sigma$ is the number of world-lines crossing $d\Sigma$.

This number is necessarily equal to the number crossing the corresponding oblique section $dx\,dy\,dz$ in the plane $t =$ constant in A's frame. The latter number is $\varrho\,dx\,dy\,dz$. Accordingly in any system of coordinates

$$J\,d\Sigma = \varrho\,dx\,dy\,dz. \quad (116)$$

By the formula (23), § 8, Part II, we have

$$d\Sigma = dx\,dy\,dz\,\frac{c}{(c^2 - \Sigma u^2)^{1/2}}, \quad (23)$$

hence

$$J = \varrho\,\frac{(c^2 - \Sigma u^2)^{1/2}}{c}. \quad (117)$$

Now J is a definite function of position in the time-space map. It may be expressed either in terms of the co-ordinates used by A, or in terms of the co-ordinates used by B. In either case A and B will arrive at the same number J for the intensity of world-lines at a given event. Let then

$$J = \sigma\,(x, y, z, t) = \sigma_1\,(x', y', z', t'), \quad (118)$$

where

$$x' = \frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - xV/c^2}{(1 - V^2/c^2)^{1/2}}. \tag{119}$$

It then follows from (115) and (117) that

$$\sigma(x, y, z, t) = \chi(x, y, z, t) \frac{\left[c^2 - \sum_{\alpha, \beta, \gamma} \{\alpha(x, y, z, t)\}^2\right]^{1/2}}{c}, \tag{120}$$

$$\sigma_1(x', y', z', t') = \chi_1(x', y', z', t') \frac{\left[c^2 - \sum_{\alpha, \beta, \gamma} \{\alpha_1(x', y', z', t')\}^2\right]^{1/2}}{c}. \tag{120'}$$

Formula (118), (120), (120′) relate χ and χ_1, to one another.

We have further that the velocities (α, β, γ), $(\alpha_1, \beta_1, \gamma_1)$, when relating to the same particle (x, y, z) or (x', y', z') at corresponding times (t, t') are related by EINSTEIN's transformation, namely

$$\alpha_1(x', y', z', t') = \frac{\alpha(x, y, z, t) - V}{1 - \alpha(x, y, z, t)\, V/c^2}, \tag{121}$$

$$\beta_1(x', y', z', t') = \frac{\beta(x, y, z, t)(1 - V^2/c^2)^{1/2}}{1 - \alpha(x, y, z, t)\, V/c^2}, \tag{122}$$

$$\gamma_1(x', y', z', t') = \frac{\gamma(x, y, z, t)(1 - V^2/c^2)^{1/2}}{1 - \alpha(x, y, z, t)\, V/c^2}. \tag{123}$$

77. So far we have only used the first postulate of relativity, the constancy of the velocity of light; we have simply used the LORENTZ formula to render the two observers' decriptions of the flow consistent.

We now apply the second postulate — that if the flow is to represent the world, the descriptions must not only be consistent but also identical. This demands

$$\chi \equiv \chi_1, \tag{124}$$

$$\alpha \equiv \alpha_1, \quad \beta \equiv \beta_1, \quad \gamma \equiv \gamma_1. \tag{124'}$$

Equations (120), (120′) now show that

$$\sigma \equiv \sigma_1.$$

Introducing this in (118) we have, taking into account (119),

$$\sigma\left(\frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, \; y, \; z, \; \frac{t - xV/c^2}{(1 - V^2/c^2)^{1/2}}\right) = \sigma(x, y, z, t) \tag{125}$$

a functional equation for σ which has to be satisfied for all values of x, y, z, t independently of V. Two other functional equations are obtained by letting B move along the y- and z-axes respectively.

78. By the method previously employed, it is readily shown that the general solution of (125) and the two similar equations is

$$\sigma = \sigma(X), \tag{126}$$

where as usual

$$X = t^2 - \frac{\Sigma x^2}{c^2}.$$

Now σ is of dimensions -3 in length. Hence the only possible form of σ is

$$\sigma = \frac{D}{c^3 X^{3/2}}, \tag{127}$$

where D is a constant which is a pure number.

As this argument may not appear convincing I shall give an independent proof of (127) below, para 81.

79. Identities (124′) introduced in (121), (122), (123) give the functional equations

$$\alpha\left(\frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, y, z, \frac{t - Vx/c^2}{(1 - V^2/c^2)^{1/2}}\right) = \frac{\alpha(x, y, z, t) - V}{1 - \alpha(x, y, z, t)V/c^2}, \tag{128}$$

$$\beta\left(\frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, y, z, \frac{t - Vx/c^2}{(1 - V^2/c^2)^{1/2}}\right) = \frac{\beta(x, y, z, t)(1 - V^2/c^2)^{1/2}}{1 - \alpha(x, y, z, t)V/c^2}, \tag{129}$$

$$\gamma\left(\frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, y, z, \frac{t - Vx/c^2}{(1 - V^2/c^2)^{1/2}}\right) = \frac{\gamma(x, y, z, t)(1 - V^2/c^2)^{1/2}}{1 - \alpha(x, y, z, t)V/c^2}. \tag{130}$$

We have also six further similar equations obtained by letting B move along the y- and z-axes. These additional equations may be replaced by the statement that the solution of (128), (129), (130) is to be a vector in 3-space.

The general solution of (128) is easily found to be

$$\alpha(x, y, z, t) = c\,\frac{(ct + x) + (ct - x)\,\Psi(c^2t^2 - x^2; y, z)}{(ct + x) - (ct - x)\,\Psi(c^2t^2 - x^2; y, z)}, \tag{131}$$

where Ψ is arbitrary. By the solution of the two similar equations for β and γ we must have also

$$\beta(x, y, z, t) = c\,\frac{(ct + y) + (ct - y)\,\Psi(c^2t^2 - y^2; z, x)}{(ct + y) - (ct - y)\,\Psi(c^2t^2 - y^2; z, x)}, \tag{131′}$$

$$\gamma(x, y, z, t) = c\,\frac{(ct + z) + (ct - z)\,\Psi(c^2t^2 - z^2; x, y)}{(ct + z) - (ct - z)\,\Psi(c^2t^2 - z^2; x, y)}. \tag{131″}$$

These must satisfy the further equations (129), (130) and their analogues; or alternatively (α, β, γ) must be a 3-vector. The resulting algebra I have found cumbersome, and I do not claim to have constructed a rigorous proof, but the only function Ψ I can find consistent with all the conditions is

$$\Psi(c^2t^2 - x^2;\ y, z) \equiv \Psi(c^2t^2 - x^2 - y^2 - z^2) \equiv -1. \tag{132}$$

This gives, inserted in (131), (131′), (131″),

$$\alpha(x, y, z, t) = \frac{x}{t}, \quad \beta(x, y, z, t) = \frac{y}{t}, \quad \gamma(x, y, z, t) = \frac{z}{t}, \tag{133}$$

which is an exact solution of the problem and presumably the only one. It determines the velocity as a function of position and time. It follows that

$$\frac{[c^2 - \sum_{\alpha, \beta, \gamma} \{\alpha(x, y, z, t)\}^2]^{1/2}}{c} = \frac{X^{1/2}}{t},$$

whence by (120) and (127)

$$\chi(x, y, z, t) = \frac{D\,t}{c^3 X^2} = \frac{D\,t}{c^3 (t^2 - \Sigma x^2/c^2)^2} \tag{134}$$

which determines the density distribution.

80. It is now found that with

$$\varrho = \frac{D\,t}{c^3 X^2}, \quad u = \frac{x}{t}, \quad v = \frac{y}{t}, \quad w = \frac{z}{t} \tag{135}$$

the equation of continuity (114) is satisfied identically. Lastly the x-component of acceleration of an individual fluid element is Du/Dt[1]) where

$$\left.\begin{aligned} \frac{Du}{Dt} &= \left[\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}\right]\frac{x}{t} \\ &\equiv 0 \end{aligned}\right\} \tag{136}$$

and so every particle is in uniform rectilinear motion.

81. The dimensional argument by which we evaluated σ would break down if an absolute natural constant existed of the dimensions of a time[2]). We now show that this is impossible, on our postulates. Take the solution $u = x/t$, $v = y/t$, $w = z/t$, $\sigma = \sigma(X)$, σ being supposed unknown. This gives

$$\varrho = \chi(x, y, z, t) = \frac{c\,\sigma(X)}{(c^2 - \Sigma u^2)^{1/2}} = \frac{t\,\sigma(X)}{X^{1/2}}$$

Introduce these values for ϱ, u, v, w in the equation of continuity (114). We get

$$\frac{\partial}{\partial t}\left[\frac{t\,\sigma(X)}{X^{1/2}}\right] + \sum \frac{\partial}{\partial x}\left[\frac{t\,\sigma(X)}{X^{1/2}}\,\frac{x}{t}\right] = 0$$

which after some algebra reduces to

$$3\,\sigma(X) + 2\,X\,\sigma'(X) = 0. \tag{137}$$

1) D/Dt denotes the Eulerian operator "following the motion".

2) Or, alternatively, using c, of the dimensions of a length.

The fact that X alone occurs in this equation is a check on our velocity-formulae. The solution of (137) is

$$\sigma(X) = \frac{D}{c^3 X^{3/2}}$$

in agreement with (127), D being of zero dimensions.

82. We have now found that the only possible flow is that of centrally symmetrical expansion

$$u = x/t, \quad v = y/t, \quad w = z/t \tag{138}$$

with a density law (number of particles per cm.³)

$$\varrho = \frac{D\,t}{c^3\,(t^2 - \Sigma\, x^2/c^2)^2}. \tag{139}$$

This is identical with the density-law anticipated on general grounds in § 12 and given in (104) and (105). It gives a finite density for all $t > 0$ (diminishing with increasing t) at all points (x, y, z) within the open expanding sphere

$$x^2 + y^2 + z^2 < c^2 t^2.$$

There is an exact correlation between velocity and position of precisely Hubble's type; and the density becomes indefinitely great as observations are made at greater and greater distances within the sphere. This description holds for any observer whatever, situated on any particle taking part in the flow. Thus if an observer leaves an assigned particle N_1, regarded as the centre of the universe, moves so as to overtake a second nebula N_2, and then moves along with N_2, he will still see himself as the centre of the nebular system of flow, and he will observe the same density-distribution (in his new time) as before. In the space framework of any one observer, the world of nebulae is finite in extent at any instant. But it is infinite in the sense that if a second observer leaves him with any arbitrary velocity, he can go on for ever without coming to the boundary of the system. We reconcile our antipathy to a system filling infinite Euclidean space with our sympathy with the *concept* of infinite Euclidean space.

We can now see why every particle of the fluid, i. e. every nebula, undergoes no acceleration. The observer moving with any assigned nebula sees himself as the centre of the system of nebulae, distributed with spherical symmetry round himself. Hence his nebula must necessarily be unaccelerated. But the same is true of any other observer on any other nebula. Hence that also is unaccelerated. Hence the nebulae can only be in uniform rectilinear motion whatever their inter-actions, i. e. whatever the law of gravitation.

83. This appears at first sight to be in contradiction with our ordinary ideas of Newtonian gravitation. If N_0 is the centre of the system (to one observer), N_1 another nebula, eccentric to him, then we are inclined to say that N_1, must be subject to the gravitational acceleration of all the nebulae lying inside the sphere centre N_0 and radius N_0N_1; for by a theorem due to NEWTON, the spherical shell between N_1 and the limits of the system might be expected to have zero effect on particles inside it. But a moment's reflection shows that such an application of NEWTON's theorem would be unjustifiable, for its method of proof holds only for a finite shell at rest. Without very careful consideration we are not entitled to apply it to the outer thin shells moving with velocities approaching that of light, for it is quite ambiguous, until we have fixed the observer, what is meant by the centre of such a shell. When we take axes moving with the nebula N_1 concerned, it at once becomes the centre of the whole system, and the acceleration vanishes. Further the proof of NEWTON's theorem fails when the density tends to infinity at some given spherical shell, for the integral concerned is then only semi-convergent. Every nebula is thus unaccelerated in its own frame, and so in all frames. I feel that these considerations and this final picture brings into clear relief, from the most elementary standpoint, a connection between light and gravitation without any appeal to a "geometrisation of physics".

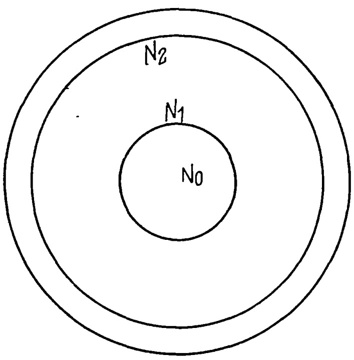

Fig. 4.

We can put the matter otherwise. In the frame of an observer O_0 on a nebula N_0, dissect the world into three portions: (1) a sphere centre N_0, radius N_0N_1; (2) a shell from N_1 to N_2, which is a nebula near the outer boundary but still moving with a velocity small compared with c; (3) the outer shell extending from N_2 to $r = ct$. Nebula N_0 is unaccelerated; hence N_1 must be unaccelerated. But N_1 is exposed to the sum of the gravitational influences of (1), (2) and (3). The effect of (2) is zero, by NEWTON's theorem. The effect of (1) is an attraction along N_1N_0. Hence the effect of (3) must be an equal and opposite attraction along N_0N_1. Did NEWTON's law *force* $\propto 1/(\text{distance})^2$ hold independently of velocity, this would scarcely be true. But there must be corrections to NEWTON's law depending on velocity and becoming large as (velocity) $\rightarrow c$, which in connection with the fact that $\rho \rightarrow \infty$ as $r \rightarrow ct$, results in the outermost region (3) having a finite radially outward attraction at any eccentric point. The further examination of this should give great insight into the nature of gravitation.

The system of radially outward flow, given by (138) and (139), the same observationally to all observers moving with the flow, wherever they be, has many remarkable properties, which I hope to analyse on another occasion. That such a system exists, and is capable of description in the Euclidean space of any arbitrary observer moving with the flow may be taken as the principal result of this paper — a result independent of course of the line of argument which has led to its discovery.

§ 14. Conclusion. 84. We have now arrived at a picture of the universe which, I believe, goes a great way towards explaining the fundamental problem concerning the location of any observer's world in time and space, and which promises ultimately to throw great light on the nature of gravitation. We have done so employing only the ordinary Euclidean space of our experience, and the ordinary clocks by which we keep our time; we have avoided any "geometrisation of physics"; as physical principles we have used only (a) Einstein's induction from those many experiments which failed to determine absolute motion, namely the constancy of the velocity of light as embodied in the Lorentz formulae of transformation, (b) Einstein's principle that all points in the world must be equivalent, significantly altered however so as to read that the evolving series of world-views seen by any one observer must be identical with the evolving series of world-views seen by any other observer. We have used no metric of space-time in any physical sense, employed no field equations or dynamical equations of motion. We have used the space-time of Minkowski not as representing any supposed real entity "the space-time of the world", but only as a convenient map used for simplifying certain calculations.

Our line of thought may be briefly recalled. We saw first that expansion is an inevitable phenomenon for an un-enclosed system in infinite space. It has nothing to do with gravitation. It occurs in the absence of gravitation, and is merely retarded by gravitation. This gave a general insight into the observed expansion phenomenon of the universe of spiral nebulae, and explained the observed velocity-distance proportionality. We were thus led to study the abstract properties of a kinematic system of particles in uniform motion and we inferred the existence (a) of a natural time-origin in the time-scale and a space-origin in the space-frame, of any observer (b) of a natural origin in any convenient space-time map. We then saw, on general grounds, that the only world, which could possibly exist would possess such a property, namely, its map in space-time would necessarily have spherical symmetry about a point in the map. Absolute position in space however then lost all meaning, and an apparent location of any observer's universe, *as observed*, in his space, was shown to be simply an effect created by the presence of the observer.

The question then arose whether it was possible further to particularise the distribution of matter and motion to be expected in the world. It immediately suggested itself that the only distribution that could be expected as constituting the world must be one which is the same for all observers, for each observer in his own time scale and space-frame[1]). We then showed by detailed mathematics that if any two observers setting out from one another at a given instant $t = 0$ were to see, unfolding before them, identical "perspectives" of the universe (each in his own time and space) then the universe of matter and motion observed by either must necessarily be of the expanding type, possess spherical symmetry in space and possess also a natural time-origin coinciding with $t = 0$. We showed further that *all* observers inside the sphere $r = ct$ (at any instant t in the time-scale of any given observer), wherever they are in the space-frame of the given observer, would see identical perspectives of the universe provided they assumed suitable, observationally ascertainable, velocities.

We were then able to show, without any appeal to dynamics, that if the two postulates of relativity held good exactly all particles in the world would describe straight lines with uniform velocity and the resultant gravitational acceleration at each point should be accordingly zero; but that singularities would be present. Reasons were then given for supposing that if one of the postulates were given up, the singularities would be removed, and any consequential accelerations could be attributed to what we ordinarily call gravitation. But in this mode of derivation, the law of gravitation is expressed neither as a set of field-equations nor as an effect of action at a distance, but simply in the statement that such and such a distribution of matter and motion in the world co-exists with such and such a distribution of accelerations. If it is preferred to analyse this result either into field equations or into action at a distance, this would be a legitimate mode of procedure; but it is equivalent to introducing causal concepts which play no part in the train of thought of the present paper.

The foregoing analysis related to the possible *statistical* distributions of matter and motion in the world, i. e. distributions in which, at a given position at a given time, there exists not a definite velocity but a velocity distribution. The essential characteristic of the velocity-distributions

[1]) As my article in "Nature" (July 2, 1932) showed, my formulation of this principle was made independently of any knowledge of Einstein's less detailed formulation of a similar though far from identical principle.

thus determined was the existence of velocities ranging up to c. We briefly outlined the probable modifications that would be introduced by Newtonian gravitation on such kinematic systems, and saw reason to suppose that there would be formed an endless sequence of gravitationally closed, similar sub-systems, escaping from one another, lying inside the sphere $r = ct$. We were led to expect that these sub-systems would occupy the positions of the singularities (regarded as density-maxima) of the original statistical system. The space-velocity distribution of the sub-systems could thus be inferred, and it was found to give a strong concentration towards the inner boundary of the sphere $r = ct$, each sub-system receding from a "common centre" with a velocity-distance proportionality, the "common-centre" being any arbitrary observer situated on our arbitrarily chosen sub-system.

Such a system of sub-systems was then seen to exhibit not a *statistical* velocity-distribution at each point but an exact correlation between velocity and position, and therefore to exhibit the characteristics of a *hydrodynamical* motion of flow. We therefore investigated the most general possible hydrodynamical motion of three-dimensional flow compatible with the two postulates of relativity, again without any appeal to dynamics. We found[1]) that the most general motion of flow was precisely that anticipated on general grounds, namely a radial outward motion and a spatial distribution (with correlated velocity-distribution) coinciding with the motions of the singularities in the previously considered statistical system. The hydrodynamical solution is free from singularities save on the sphere $r = ct$. In the hydrodynamical solution every particle is free from acceleration and the resultant gravitational acceleration at any point is accordingly zero. We saw in a general way how this result was not incompatible with the known theorems on Newtonian gravitation.

Relativity — statistics — hydrodynamics — dynamics — gravitation — such was our course of investigation. The final step[2]) remaining to be accomplished is the discussion of a relativistically invariant system of *discrete* masses — the next step after statistical motion and hydrodynamical flow. I anticipate that all sub-systems will then be found to be points where the gravitational acceleration is rigorously zero. The final law of gravitation may be contained in the statement that such a relativistically invariant system of discrete point-masses is possible, and that the self-gravitation of the complete system is such that each point-mass is found

[1]) Apart from from one step of imperfect mathematical rigour.

[2]) I have begun this investigation but considerations of space compel me to postpone the results to another occasion.

at a zero of the gravitational acceleration. Such a possible formulation of the law of gravitation is to be distinguished in a marked way from previous formulations in terms of action at a distance or field equations.

We finally identify the grand system of closed sub-systems with the universe as a whole, and the separate sub-systems with spiral nebulae. These nebulae *appear* to have separated from or issued out of a limited region in the vicinity of any arbitrary observer, moving with an arbitrarily selected nebula, at the instant $t = 0$ in his time-scale; though extrapolation to $t = 0$ is essentially unjustifiable as the radius of the universe of observable events is then zero and no observations are possible. The mean density of the universe decreases at any point as the time increases. The *number* of nebulae per unit volume, *near the observer* decreases as the inverse cube of the time; the product of the so-called "constant" of gravitation and the mean density of the smoothed-out universe near the observer, in grams, decreases as the inverse square of the time[1]). The upper limit found for this mean density, 10^{-27} gram · cm^{-3} is consistent with observation.

Observation is confined to the interior of the sphere $r = ct$. The sub-systems are strongly concentrated towards the inner boundary of this expanding sphere, and become infinitely numerous as the boundary is approached, every point of the expanding sphere being a limiting point. But the light received from a single spiral nebula, moving with recession-velocity close to c, near this boundary is reduced by the recession, and in such a way as to give the *appearance* of a spatially uniform distribution of nebulae, i. e. a distribution apparently uniform when the distance is (erroneously) judged by brightness alone without correction for recession-velocity. The total light received by an observer or a photographic plate at the origin is finite, although the number of nebulae is infinite. "What exists" outside the sphere $r = ct$ is not a scientific question, since observations beginning at $t = 0$ can never extend further than $r = ct$.

Thus in some way we may suppose the relativistic statistical particle distribution to be the primeval ground-plan of the universe and to give rise to a system of receding nebulae. The invariant properties both of the primeval distribution of matter and motion and of the consequential system of nebular sub-systems go far to solve the problems of the location of this world of ours in space and time, for they show that an apparent

---

[1]) It follows from these relations that the "constant of gravitation G" is not a constant but is proportional to the time t measured from the natural time-origin. This is so revolutionary a suggestion that I postpone the discussion of it and a host of allied results to a separate paper.

"whereness" of the world in space and "whenness" of the world in time are simply effects produced by the (necessary) introduction of an observer. The world is three-dimensional. It has no centre in "Time" or in "Space", only a centre in the space and an apparent beginning in the time of any arbitrarily chosen observer.

This primeval ground plan of matter and motion and this consequential system of sub-systems, exhibit an unfolding panorama of expansion, dilution and velocity-segregation, the same for every arbitrary observer. Its other properties — its evolutionary significance, its fixation of a natural frame of zero rotation, its compatibility with complete observational knowledge of every observer's past picture and complete ignorance of every observer's future picture — I shall deal with elsewhere.

85. In view of certain lacunae in my work, I do not wish to press the details of the gravitational state of the world as a whole. I see no alternative set of possibilities; the expanding sphere of receding, infinitely numerous nebular systems, each the observational centre of all the others, is so satisfying philosophically, physically and observationally that I feel that the law of gravitation must be such as to be compatible with it. But I wish chiefly to lay stress on the new methods of investigation here developed, and those mathematical results which have been established rigorously. I wish to stress also that these new methods have arisen from simple physical, almost common-sense considerations. These considerations show that the expansion phenomenon has nothing to do with gravitation in the first instance. They further show that it is not necessary to employ the vague, ill-understood, probably meaningless concept of closed, finite expanding space, but that simple Euclidean space suffices for all purposes. Lastly the difficulties which led Einstein to suggest a spherical form for the world — difficulties of a definite physical character — have been found to disappear on further analysis.

It will be apparent to every reader that many other world-problems can be tackled by the methods of this paper. Part III contains an example.

III. Light.

1. We saw in part II of this paper that the spatio-velocity distribution of the particles of the statistical system subject only to the two postulates of relativity was

$$f(x, y, z, t, u, v, w)\, dx\, dy\, dz\, du\, dv\, dw = \frac{C\, dx\, dy\, dz\, du\, dv\, dw}{c^6\, Y\, (Z^2 - X\, Y)^{3/2}} \qquad (1)$$

with

$$X = t^2 - \frac{\Sigma x^2}{c^2}, \tag{2}$$

$$Y = 1 - \frac{\Sigma u^2}{c^2}, \tag{3}$$

$$Z = t - \frac{\Sigma u x}{c^2}. \tag{4}$$

This formula clearly breaks down when $u^2 + v^2 + w^2 = c^2$. Not only so, but the method of derivation breaks down, for it is impossible to describe a velocity-element $du\,dv\,dw$ about the velocity $|u, v, w| = c$, without going to velocities exceeding c. Thus reconsideration is necessary.

When $|u, v, w| = c$ we have $Y = 0$. Hence the only invariant that can appear in the specification of particles moving with velocity is

$$X = t^2 - \frac{\Sigma x^2}{c^2} \tag{5}$$

and the only covariant is

$$W = t - \frac{\Sigma l x}{c} \tag{6}$$

where $l : m : n$ is the direction of motion and $l^2 + m^2 + n^2 = 1$. It is however impossible to construct a zero-dimensional invariant out of these expressions together with c; for the only non-dimensional combination is W^2/X, and this is not invariant under a Lorentz transformation, since W is only covariant. Yet we require a zero-dimensional invariant as the argument of the distribution function.

Again, no invariant expression of the type

$$f(x, y, z, t, l, m, n)\, dx\, dy\, dz\, d\omega$$

for the number of particles with velocity c inside $dx\,dy\,dz$, in the cone $d\omega$, at time t, can be constructed out of W, X and c. It is clear that to represent these particles at all, some further parameter is required descriptive of the particles. This parameter must be such as, in combination with W, X and c, will permit the construction of an argument of zero dimensions invariant under a Lorentz transformation. It is clear from consideration of (5) and (6), (1) that the new parameter must be of dimensions either of a time or a length (and by means of c these can be converted into one another); (2) that the new parameter must be such a to make with W a Lorentz invariant (for X is already a Lorentz invariant).

2. For definiteness I shall therefore adopt as the new parameter a variable ν, having the dimensions of an inverse time.

$$\nu W \tag{7}$$

is to be a Lorentz invariant.

The particles will be each characterized by a definite value of the parameter ν. They will therefore have a distribution in ν. Hence we must seek the number lying inside $(\nu, \nu + d\nu)$. All such particles move precisely with the velocity c, and therefore we cannot construct a velocity-range $du\,dv\,dw$ round c. Instead we construct a direction range $d\omega$ round an assigned direction l, m, n, in three dimensions. Lastly, we can either count the number (lying inside $d\nu\,d\omega$) lying also inside $dx\,dy\,dz$ at time t, or we can count the number crossing a small element of surface dS perpendicular to l, m, n, during an interval dt. The former method will specify their *density*-distribution, the latter their *intensity*-distribution. I have found it easier and more instructive to adopt the latter.

3. I therefore let

$$N(\nu, t, x, y, z, l, m, n)\,d\nu\,d\omega\,dt\,(i \cdot dS) \tag{8}$$

be the number of particles of velocity c crossing the element of surface dS normal to i (l, m, n), in time dt, inside $d\nu\,d\omega$ at the point (x, y, z) at the instant t, in the frame of a given observer A. Similarly, let

$$N_1(\nu', t', x', y', z', l', m', n')\,d\nu'\,d\omega'\,dt'\,(i' \cdot dS'), \tag{9}$$

be the distribution law observed by a second observer B moving with velocity V with respect to the first observer A along A's x-axis, where $\nu', \ldots n', d\nu', \ldots dS'$ are measured in B's own frame.

By the Lorentz transformation, if the two observers part company at time $t = 0$ from an origin $x = 0$, $y = 0$, $z = 0$, a particle in direction l, m, n at x, y, z, t in A's reckoning is observed to have direction l', m', n', and to be at x', y', z', t', by B, in B's reckoning, where

$$x' = \frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - Vx/c^2}{(1 - V^2/c^2)^{1/2}}, \tag{10}$$

$$l' = \frac{l - V/c}{1 - lV/c}, \quad m' = \frac{m(1 - V^2/c^2)^{1/2}}{1 - lV/c}, \quad n' = \frac{n(1 - V^2/c^2)^{1/2}}{1 - lV/c}. \tag{11}$$

From these

$$W' = W\,\frac{(1 - V^2/c^2)^{1/2}}{1 - lV/c}. \tag{12}$$

Hence since νW is to be a Lorentz invariant,

$$\nu' = \nu\,\frac{1 - lV/c}{(1 - V^2/c^2)^{1/2}}, \tag{13}$$

$$d\nu' = d\nu\,\frac{1 - lV/c}{(1 - V^2/c^2)^{1/2}}. \tag{14}$$

Further

$$dt' = \frac{dt - V\,dx/c^2}{(1 - V^2/c^2)^{1/2}}$$

in which

$$dx = V\,dt,$$

so that

$$dt' = dt\,(1 - V^2/c^2)^{1/2}. \tag{15}$$

Again, by a projection argument

$$\frac{d\omega'}{d\omega} = \frac{l}{l'}\frac{dm'\,dn'}{dm\,dn} = \frac{l}{l'}\left[\frac{\partial(m', n')}{\partial(m, n)}\right]_{l=(1-m^2-n^2)^{1/2}} = \frac{1 - V^2/c^2}{(1 - l\,V/c)^2}$$

on using (11). Lastly the components of dS, namely dS_x, dS_y, dS_z in A's reckoning, are related to $dS'_{x'}$, $dS'_{y'}$, $dS'_{z'}$ by

$$dS'_{x'} = dS_x, \quad dS'_{y'} = \frac{dS_y}{(1 - V^2/c^2)^{1/2}}, \quad dS'_{z'} = \frac{dS_z}{(1 - V^2/c^2)^{1/2}}, \tag{16}$$

so that

$$(l' \cdot dS') = \Sigma\, l'\, dS'_{x'} = \frac{(l - V/c)\,dS_x + m\,dS_y + n\,dS_z}{1 - l\,V/c}. \tag{17}$$

4. The observer A with the stationary element dS observes the number of particles (8). B, with an element dS' stationary in his own frame, observes the number as (9) and informs A of this. A sees the moving element dS' as receding from the radiation incident from behind it. A's object is to equate the number of particles he actually observes to the number B says he (B) observes, corrected for B's motion. Hence (8) must equal the sum of (9) and

$$\frac{N}{c} V\,dt\,d\omega\,d\nu\,dS_x \tag{18}$$

the latter being the number which B's element has missed owing to its forward motion with velocity V in A's frame; (18) is of course calculated in A's frame by A himself from his own observations. It follows that

$$\begin{aligned} N(\nu, t, x, y, z, l, m, n)\,d\nu\,d\omega\,dt\,(l\,dS_x + m\,dS_y + n\,dS_z) \\ = N_1(\nu', t', x', y', z', l', m', n') \\ \cdot d\nu \frac{1 - l\,V/c}{(1 - V^2/c^2)^{1/2}} \cdot d\omega \frac{1 - V^2/c^2}{(1 - l\,V/c^2)^2} \cdot dt\,(1 - V^2/c^2)^{1/2} \\ \cdot \frac{(l - V/c)\,dS_x + m\,dS_y + n\,dS_z}{1 - l\,V/c} + \frac{N}{c} V\,dt\,d\omega\,d\nu\,dS_x. \end{aligned} \tag{19}$$

Equating to zero the coefficients of dS_x, dS_y, dS_z independently we get the single relation

$$\begin{aligned} N(\nu, t, x, y, z, l, m, n) = \frac{1 - V^2/c^2}{(1 - l\,V/c^2)^2} N_1\Big(\nu \frac{1 - l\,V/c}{(1 - V^2/c^2)^{1/2}}, \\ \frac{t - V\,x/c^2}{(1 - V^2/c^2)^{1/2}}, \frac{x - V\,t}{(1 - V^2/c^2)^{1/2}}, y, z, \ldots, \ldots, \frac{n\,(1 - V^2/c^2)^{1/2}}{1 - l\,V/c^2}\Big). \end{aligned} \tag{20}$$

5. Equation (20) embodies the principle of special relativity. If we now impose the postulate of the extended principle of relativity, namely that the distribution of particles must appear the same to all observers, we have further identically

$$N \equiv N_1. \tag{21}$$

Equation (20) then gives a functional equation for N which must be satisfied identically in v, t, x, y, z, l, m, n for all V. Two other similar equations can be obtained by considering B as moving along the y- and z-axes.

6. To solve equations of the type (20), expand to the first power of V in accordance with previous procedure (see Part II) and equate to zero the first power of V. We find

$$2lN - vl\frac{\partial N}{\partial v} - \frac{x}{c}\frac{\partial N}{\partial t} - ct\frac{\partial N}{\partial x} - (1-l^2)\frac{\partial N}{\partial l} + lm\frac{\partial N}{\partial m} + ln\frac{\partial N}{\partial n} = 0. \tag{22}$$

Some care is required in the meanings of $\partial N/\partial l$, etc., since l, m, n are not independent. We will therefore suppose that N has been made homogeneous and of zero dimensions in l, m, n. We have further two similar equations obtained by cyclical interchange of l, m, n and x, y, z. We seek a solution of the type

$$F(N, v, t, x, y, z, l, m, n) = 0 \tag{23}$$

and write equations (22) in the forms

$$D_1 F = 0, \quad D_2 F = 0, \quad D_3 F = 0,$$

where

$$D_1 \equiv -2lN\frac{\partial}{\partial N} - vl\frac{\partial}{\partial v} - \frac{x}{c}\frac{\partial}{\partial t} - ct\frac{\partial}{\partial x} - (m^2+n^2)\frac{\partial}{\partial l} + lm\frac{\partial}{\partial m} + ln\frac{\partial}{\partial n},$$

etc. We then find that

$$(D_2 D_3 - D_3 D_2)F = \left[y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} - (l^2+m^2+n^2)\left(n\frac{\partial}{\partial m} - m\frac{\partial}{\partial n}\right)\right]F = 0 \tag{24}$$

and two similar relations. Exactly as in Part II, § 8, it may be shown that the six equations (22) and (24) are equivalent to five independent equations and form a "complete" set. The number of independent variables $N, v, t, x, y, z, l, m, n$ being nine, it follows that F is a function of (9—5) or 4 independent integrals of the equations. These are readily found to be

$$\frac{N}{v^2}, \quad t^2 - \frac{\Sigma x^2}{c^2}, \quad v\left(t - \frac{\Sigma l x}{c}\right), \quad l^2+m^2+n^2.$$

It follows from (23) that the general solution is of the form

$$N = \nu^2 \Phi (X, \nu W) \tag{25}$$

and that the distribution law is accordingly

$$N(\nu, t, x, y, z, l, m, n)\, d\nu\, d\omega\, dt\, (i \cdot dS) = \frac{\nu^2}{c^2} \Phi (X, \nu W)\, d\nu\, d\omega\, dt\, (i \cdot dS). \tag{26}$$

This being a pure number, it follows that Φ must be a pure number. Hence its argument must be a pure number. Hence since νW is of zero dimensions and X is of dimensions $(time)^2$, the only non-dimensional combination of νW, X and c is νW itself; X cannot appear. Hence the distribution law is

$$N(\nu, t, x, y, z, l, m, n)\, d\nu\, d\omega\, dt\, (i \cdot dS) = \frac{\nu^2}{c^2} \Phi \left(\nu\left(t - \frac{\Sigma l x}{c}\right)\right) d\nu\, d\omega\, dt\, (i \cdot dS). \tag{27}$$

This is thus the number of particles of speed c crossing dS in direction l, m, n in time dt inside $d\omega\, d\nu$. It is readily verified that (27) holds not only for small values of V but for all values of V.

7. We now have to impose the Boltzmann-equation condition — the statistical condition that (27) represents a permanent collection of *things*. We have to envisage the possibility that the particles pursue curved paths. If so, their equations of motion will be

$$\frac{dx}{dt} = cl, \quad \frac{dy}{dt} = cm, \quad \frac{dz}{dt} = cn, \tag{28}$$

$$\frac{dl}{dt} = p, \quad \frac{dm}{dt} = q, \quad \frac{dn}{dt} = r, \tag{29}$$

where p, q, r are dependent upon ν, t, x, y, z, l, m, n and subject to

$$pl + qm + rn = 0. \tag{30}$$

Consider the course of a pencil of particles. The pencil will not remain constant in elementary solid angle. All the particles passing through dS inside $d\omega$ in time dt must also pass through dS' inside dS' in time dt, where dS' is a later cross-section of the pencil and $d\omega'$ is the later elementary solid angle of the pencil[1]). It follows, since $dS' = dS$ to a sufficient order, that

$$N(\nu, t, x, y, z, l, m, n)\, d\omega = N(\nu, t + \Delta t, x + \Delta x, \ldots, \ldots, l + \Delta l, \ldots, \ldots)\, d\omega',$$

[1]) These symbols dS', $d\omega'$ must not be confused with the previous symbols dS', $d\omega'$ for the relativistic transforms. Here dS' and $d\omega'$ refer to a later position along one and the same pencil.

where
$$\Delta x = c l \Delta t, \quad \Delta y = c m \Delta t, \quad \Delta z = c n \Delta t,$$
$$\Delta l = p \Delta t, \quad \Delta m = q \Delta t, \quad \Delta n = r \Delta t$$
and
$$d\omega' = d\omega \left(1 + \Delta t \sum \frac{\partial p}{\partial l}\right) d\omega\ ^{1)}. \tag{31}$$
It follows that N must satisfy
$$\frac{\partial N}{\partial t} + c \sum l \frac{\partial N}{\partial x} + \sum p \frac{\partial N}{\partial l} + N \sum \frac{\partial p}{\partial l} = 0. \tag{32}$$
Inserting for N from (27), and noting that
$$\frac{\partial \Phi}{\partial t} + c \sum l \frac{\partial \Phi}{\partial x} \equiv 0 \tag{33}$$
we have as our condition
$$\sum p \frac{\partial \Phi}{\partial l} + \Phi \sum \frac{\partial p}{\partial l} = 0$$
or
$$\sum \frac{\partial}{\partial l} (p \Phi) = 0. \tag{34}$$
This is the relation which the curvature functions p, q, r must satisfy.

8. We saw in Part II that the accelerations of particles moving with a velocity u, v, w of modulus less than c were of the form
$$p = \frac{Y}{X} (x - u t) F\left(\frac{Z^2}{X Y}\right)$$
etc. We require the limiting form of such expressions as $u^2 + v^2 + w^2 \to c^2$. The limits of p, q, r must necessarily be finite as $u^2 + v^2 + w^2 \to c^2$. Now as $u^2 + v^2 + w^2 \to c^2$, we have $Y \to 0$ and $\xi = Z^2/XY \to \infty$. I shall assume that either $F(\xi)$ has a finite limit as $\xi \to \infty$ or $F(\xi) \to \infty$. In the former case since $Y \to 0$ we have $p, q, r \to 0$ as $u^2 + v^2 + w^2 \to c^2$. In the latter case, for p, q, r to remain finite, we must have $F(\xi) \sim D\xi$ as $\xi \to \infty$, where D is some constant[2]). In that case
$$p = D \lim \frac{Y}{X} \cdot \frac{Z^2}{X Y} (x - u t)$$
$$= D (x - c l t) \frac{W^2}{X^2} \cdot$$
But inserting in (30) we find
$$D (\Sigma l x - c t) \frac{W^2}{X^2} = 0$$
or
$$D W^3 / X^2 = 0$$

---

[1]) I omit this calculation, which is somewhat long. p, q, r are supposed to have been rendered homogeneous of degree zero in l, m, n before differentiations are performed.

[2]) If $F(\xi)$ tends to infinity less rapidly than $D\xi$ the same conclusion of course follows.

whence
$$D = 0. \tag{35}$$
Hence
$$p,\ q,\ r = 0 \tag{36}$$
and l, m, n are constant along any world-line. Hence the tracks are straight lines pursued with velocity c and (34) is identically satisfied.

9. Thus finally the distribution of intensity, of particles moving with velocity c, in time, space and direction of motion, is given by (27), with l, m, n constant along the path of any particle.

10. It is now clear that what we have called particles of velocity c characterised by a *parameter* ν are indistinguishable from photons of frequency ν.

This means that the state of the world implied by the two postulates of relativity includes *light* amongst its constituents, moreover light characterised by a frequency distribution. This is necessary for consistency, as we have already postulated the existence of light in using the Lorentz formula which are based on signalling experiments. The analysis now shows that the light used in signalling has exactly the properties of *particles of the system* moving with speed c. We have further shown that the particle-paths are straight lines, in agreement with the first postulate; for were the lighttracks curved, the Lorentz formula would be altered. That is, the two postulates of relativity imply that light is not subject to gravitation, as already inferred.

I have not succeeded in further determining the function Φ occurring in (27). If it is really arbitrary, we have established a strong disjunction between light and matter, for we were able fully to determine the distribution of matter in motion with speeds *less than* c. All that we can infer from (27) is the existence of independent plane waves of light; for the intensity at any time t in a given direction l, m, n is constant at all points for which
$$lx + my + nz = \text{constant},$$
i. e. at all points over a plane perpendicular to the given ray. Further the intensity-distribution is isotropic in direction. But at any given point x, y, z as t alters, the change of intensity with time, in direction l, m, n is quite arbitrary for any one frequency.

Appendix. The apparent brightness of a receding nebula.

1. The object of this appendix is to calculate the change in apparent brightness of a nebula receding with velocity V due to its motion. We wish to determine the intensity of the nebula as recorded on a photographic plate placed at the origin. The problem is a little complicated because we must

measure the intensity received in units appropriate to a frame of reference in which the photographic plate is fixed, yet the intrinsic properties of the nebula — its intrinsic brightness — must be supposed known (and standard) in a frame in which the nebula is at rest. We begin with an abstract problem, namely we consider an emitting nebula at rest and the photographic plate in motion. We shall then transform the measured intensity on the plate into units appropriate to the frame of the plate, leaving however physical quantities intrinsic to the nebula as standard quantities known for a frame moving with the nebula. We have also to take care concerning the moment at which the distance between receiver and nebula is measured.

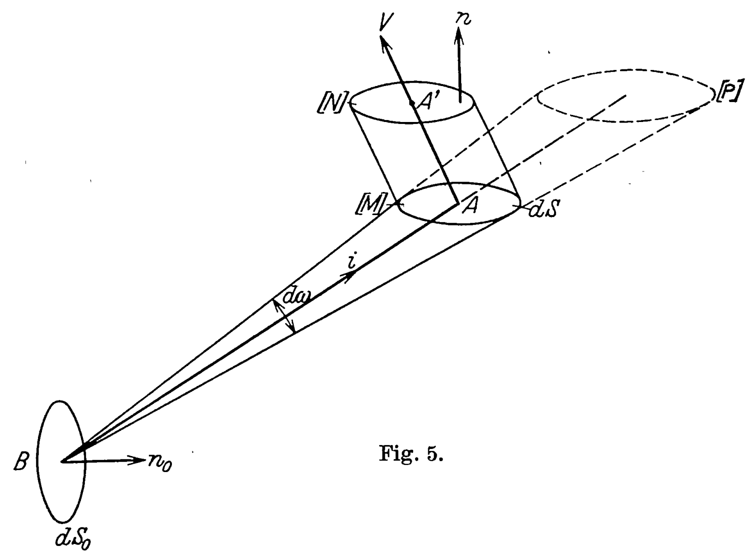

Fig. 5.

2. Consider a radiating surface dS_0 placed at a point B, with its normal in the direction of a unit vector n_0. Consider a receiving surface element dS (normal n) in motion with velocity V (a vector). Consider the radiation received by the receiver during the interval, $(t,\ t+dt)$ reckoned by a clock stationary in B's frame. Let the receiver move from A to A' during this interval. Let $BA = r$; let i be a unit vector in the direction BA, $d\omega$ the solid angle subtended by the receiver dS at B when dS itself is at A. Then

$$d\omega = \frac{d\,S\,(i \cdot n)}{r^2}. \tag{1}$$

If dS were stationary at A, the amount of energy it would receive during $(t,\ t+dt)$ would be

$$(I_\nu\, d\nu)\, d\,S_0\,(i \cdot n_0)\; d\omega\, dt = \frac{(I_\nu\, d\nu)\, dS_0\,(i \cdot n_0)\, dS\,(i \cdot n)\, dt}{r^2}. \tag{2}$$

Here I_ν is the specific intensity of radiation from dS_0 in direction i, in frequency ν, measured in a frame in which dS_0 is at rest. When dS is in motion the amount of energy received is altered. Let (M) denote the position of dS at t, (N) its position at $t + dt$. Project the area conically from B on to the plane of (N), and call this projected position (P). Then the light crossing dS as it passes from (M) to (N), to the first power of the differentials concerned, is equal to the light that would have crossed dS had it passed from (M) to (P). The error is in fact of the order of the energy passing through the curved surface of the cylinder formed by the motion of dS from (M) to (N); this energy is equal to the difference of the energies that would cross, in the same time, the area dS stationary in the positions (M) and (N) respectively, which is of a higher order than the differential we are calculating.

Now the amount falling on dS in its motion from (M) to (P) is less than the amount that would fall on dS if stationary at (M) by the amount of energy entrapped in the truncated cone cut off between (M) and (P) as subtended from B. The amount of this energy is equal to the amount of energy falling on a stationary dS at (M) in one second, namely

$$\frac{(I_\nu d\nu)\, dS_0\, (i \cdot n_0)\, ds\, (i \cdot n)}{r^2} \tag{3}$$

multiplied by the time taken by radiation to move from (M) to (P) along OA produced, namely

$$\frac{1}{c}\left[\frac{(V \cdot n)\, dt}{(i \cdot n)}\right]. \tag{4}$$

The product of (3) and (4) is

$$I_\nu\, d\nu\, dS_0\, (i \cdot n_0)\, \frac{dS}{r^2}\, \frac{(V \cdot n)\, dt}{c}. \tag{5}$$

Subtracting (5) from (2) we have for the amount of energy passing through the moving aperture

$$(I_\nu d\nu)\, dS_0\, (i \cdot n_0)\, \frac{dS}{r^2}\, dt \left[(i \cdot n) - \frac{(V \cdot n)}{c}\right] \tag{6}$$

measured in the frame of the emitting body.

3. We now consider the receiver as stationary and the emitter as moving. I shall now denote quantities measured in the moving frame, i. e. in the frame of the emitter, by accents. I shall also let A denote the observer attached to dS (the receiver), B denote the observer attached to dS_0 (the

emitter). Then the amount of energy received by the receiver dS at A, measured in B's frame, is, by the last formula,

$$(I'_{\nu'}\, d\nu')\, dS'_0\, (i' \cdot n'_0)\, \frac{dS'}{r'^2}\, dt' \left[(i' \cdot n') - \frac{(V \cdot n')}{c}\right]. \tag{7}$$

Let us simplify our problem by taking V parallel to i, of amount $V = Vi$. Then the above expression is E', say where

$$E' = (I'_{\nu'}\, d\nu')\, dS'_0\, (i' \cdot n'_0)\, \frac{dS'}{r'^2}\, dt'\, (i \cdot n') \left[1 - \frac{V}{c}\right], \tag{8}$$

since $i = i'$, each being now along V.

A will not however reckon this amount as E'. As in problems like this the quantum theory of light gives the same results as the classical wave

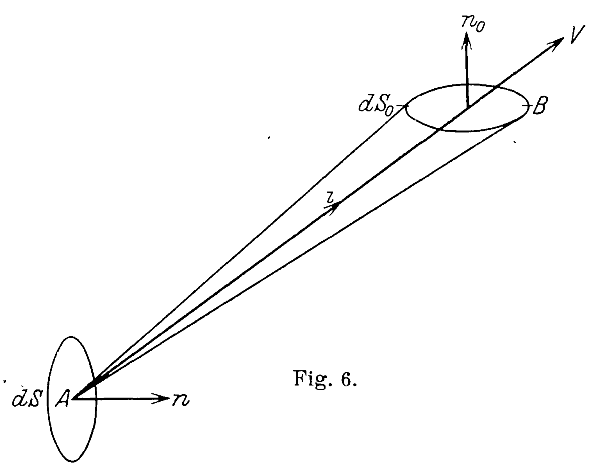

Fig. 6.

theory, it is sufficient to notice that since A receives, during the interval $(t', t' + dt')$ of B's reckoning, $E'/h\nu'$ quanta, he will receive the same number of quanta measured in his own frame, but he reckons each quantum as of amount $h\nu$, where

$$\nu = \nu' \frac{1 - lV/c}{(1 - V^2/c^2)^{\frac{1}{2}}}, \quad \text{with} \quad l = 1,$$

[see Part III, formula (13)[1])] i. e.

$$\nu = \nu' \frac{1 - V/c}{(1 - V^2/c^2)^{1/2}}. \tag{9}$$

[1]) ν and ν' are interchanged as compared with formula (13), because we are now going over from ν' to ν, not ν to ν'.

Hence he reckons the energy as E, where

$$E = \frac{E'}{h\nu'} h\nu = E' \frac{1 - V/c}{(1 - V^2/c^2)^{1/2}}. \tag{10}$$

Thus the total energy measured by A, *in terms of B's measures* is, by (8) and (10),

$$(I'_{\nu'}\, d\nu') \frac{dS'_0\,(i'\cdot n'_0)\, dS'\,(i\cdot n')\, dt'}{r'^{\,2}} \frac{(1 - V/c)^2}{(1 - V^2/c^2)^{1/2}}. \tag{11}$$

The quantity t' relates to the event which is the arrival of the light at the receiver. In B's frame (choosing an axis of x along AB) this event has co-ordinates $(-r', t')$ and in A's frame the same event is (o, t). The origin of time is so far arbitrary. For convenience adopt as origin of time the instant when B left A. This is always a possible choice, whether B was ever at A or not, for B is moving directly away from A. Then $r' = Vt'$; and the pairs (o, t) $(-Vt', t')$ describe the same event in the two frames. These pairs are now connected by the usual form of the Lorentz transformation. Accordingly

$$-Vt' = \frac{0 - Vt}{(1 - V^2/c^2)^2}, \qquad t' = \frac{t - 0}{(1 - V^2/c^2)^{1/2}}.$$

These formula are consistent; and give further

$$dt' = \frac{dt}{(1 - V^2/c^2)^{1/2}} \tag{12}$$

and

$$r' = Vt' = \frac{Vt}{(1 - V^2/c^2)^{1/2}}, \tag{13}$$

whence

$$E = (I'_{\nu'}\, d\nu') \frac{dt}{(1 - V^2/c^2)^{1/2}} \frac{dS'_0\, dS'\,(i'\cdot n')\,(i'\cdot n'_0)\cdot(1 - V^2/c^2)}{V^2 t^2} \cdot \frac{(1 - V/c)^2}{(1 - V^2/c^2)^{1/2}}, \tag{14}$$

where t is the epoch of observation in A's frame[1]).

Since dS_0 and dS are directly receding from one another and i is the direction of mutual recession, $dS'_0\,(i'\cdot n'_0)$ and $dS'\,(i'\cdot n')$, being projections normal to the direction of motion, are unaltered by a Lorentz transformation and accordingly

$$dS'_0\,(i'\cdot n'_0) = dS_0\,(i\cdot n_0),$$
$$dS'\,(i'\cdot n') = dS\,(i\cdot n);$$

[1]) It is particularly to be noted that t is not the epoch of the emission of the light but the epoch of the moment of reception.

also $i' = i$ as is easily verified and indeed is obvious. Hence the light received from dS_0 by the receiver dS and measured by the observer A at the receiver at the epoch t in his own frame is by (14)

$$E = I'_{\nu'}\, d\nu'\, dt\, dS_0\, dS\, (i \cdot n)\, (i \cdot n_0)\, \frac{(1 - V/c)^2}{V^2 t^2}\,. \tag{15}$$

We have purposely left in $\int I'_{\nu'}\, d\nu'$ in terms of B's own measures for $I'_{\nu'}\, d\nu'$ is a physical constant (the "proper" intensity of radiation of a nebula) namely a measure of the brightness of the radiating nebula measured in a frame in which it is at rest. The *total* light received by A on his own photographic plate is accordingly

$$\left(\int I'_{\nu'}\, d\nu'\right) dt\, dS_0\, dS\, (i \cdot n_0)\, (i \cdot n)\, \frac{(1 - V/c)^2}{V^2 t^2}\,, \tag{16}$$

where $(t,\ t + dt)$ gives the epoch and duration of the observation, dS and n refer to the receiver, and dS_0, n_0, V and $(\int I'_{\nu'}\, d\nu')$ refer to the radiating and moving nebula. This is the formula it was desired to obtain. The important point is the appearance of the factor $(1 - V/C)^2$, which betokens the dimming of a nebula due to recession.

Potsdam, Einstein-Institut, November 1932.