VI. *Experimental Researches in Electricity.—Seventh Series.* *By* MICHAEL FARADAY, *D.C.L. F.R.S. Fullerian Prof. Chem. Royal Institution, Corr. Memb. Royal and Imp. Acadd. of Sciences, Paris, Petersburgh, Florence, Copenhagen, Berlin, &c. &c.*

Received January 9,—Read January 23, February 6 and 13, 1834.

§. 11. *On Electro-chemical Decomposition, continued.* ¶ iv. *On some general conditions of Electro-decomposition.* ¶ v. *On a new Measurer of Volta-electricity.* ¶ vi. *On the primitive or secondary character of bodies evolved in Electro-decomposition.* ¶ vii. *On the definite nature and extent of Electro-chemical Decompositions.* §. 13. *On the absolute quantity of Electricity associated with the particles or atoms of Matter.*

Preliminary.

661. THE theory which I believe to be a true expression of the facts of electro-chemical decomposition, and which I have therefore detailed in a former series of these Researches, is so much at variance with those previously advanced, that I find the greatest difficulty in stating results, as I think, correctly, whilst limited to the use of terms which are current with a certain accepted meaning. Of this kind is the term pole, with its prefixes of positive and negative, and the attached ideas of attraction and repulsion. The general phraseology is that the positive pole *attracts* oxygen, acids, &c., or more cautiously, that it *determines* their evolution upon the surface; and that the negative pole acts in an equal manner upon hydrogen, combustibles, metals, and bases. According to my view, the determining force is *not* at the poles, but *within* the decomposing body; and the oxygen and acids are rendered at the *negative* extremity of that body, whilst hydrogen, metals, &c., are evolved at the *positive* extremity (518. 524.).

662. To avoid, therefore, confusion and circumlocution, and for the sake of greater precision of expression than I can otherwise obtain, I have deliberately considered the subject with two friends, and with their assistance and concurrence in framing them, I purpose henceforward using certain other terms, which I will now define. The poles, as they are usually called, are only the doors or ways by which the electric current passes into and out of the decomposing body (556.); and they of course, when in contact with that body, are the limits of its extent in the direction of the current. The term has been generally applied to the metal surfaces in contact with the decomposing substance; but whether philosophers generally would also apply it to the

surfaces of air (465. 471.) and water (493.), against which I have effected electro-chemical decomposition, is subject to doubt. In place of the term pole, I propose using that of *Electrode**, and I mean thereby that substance, or rather surface, whether of air, water, metal, or any other body, which bounds the extent of the decomposing matter in the direction of the electric current.

663. The surfaces at which, according to the common phraseology, the electric current enters and leaves a decomposing body, are most important places of action, and require to be distinguished apart from the poles, with which they are mostly, and the electrodes, with which they are always, in contact. Wishing for a natural standard of electric direction to which I might refer these, expressive of their difference and at the same time free from all theory, I have thought it might be found in the earth. If the magnetism of the earth be due to electric currents passing round it, the latter must be in a constant direction, which, according to present usage of speech, would be from east to west, or, which will strengthen this help to the memory, that in which the sun appears to move. If in any case of electro-decomposition we consider the decomposing body as placed so that the current passing through it shall be in the same direction, and parallel to that supposed to exist in the earth, then the surfaces at which the electricity is passing into and out of the substance would have an invariable reference, and exhibit constantly the same relations of powers. Upon this notion we purpose calling that towards the east the *anode*†, and that towards the west the *cathode*‡; and whatever changes may take place in our views of the nature of electricity and electrical action, as they must affect the natural standard referred to in the same direction, and to an equal amount with any decomposing substances to which these terms may at any time be applied, there seems no reason to expect that they will lead to confusion, or tend in any way to support false views. The *anode* is therefore that surface at which the electric current, according to our present expression, enters: it is the negative extremity of the decomposing body; is where oxygen, chlorine, acids, &c., are evolved; and is against or opposite the positive electrode. The *cathode* is that surface at which the current leaves the decomposing body, and is its positive extremity; the combustible bodies, metals, alkalies, and bases, are evolved there, and it is in contact with the negative electrode.

664. I shall have occasion in these Researches, also, to class bodies together according to certain relations derived from their electrical actions (822.); and wishing to express those relations without at the same time involving the expression of any hypothetical views, I intend using the following names and terms. Many bodies are decomposed directly by the electric current, their elements being set free; these I propose to call *electrolytes*§. Water, therefore, is an electrolyte. The bodies which,

* *ἤλεκτρον*, and *ὁδος a way*.

† *ἀνα upwards, ὁδος a way*; the way which the sun rises.

‡ *κατα downwards, ὁδος a way*; the way which the sun sets.

§ *ἤλεκτρον*, and *λυω solvo*. N. Electrolyte, V. Electrolyze.

like nitric or sulphuric acids, are decomposed in a secondary manner (752. 757.), are not included under this term. Then for *electro-chemically decomposed*, I shall often use the term *electrolyzed*, derived in the same way, and implying that the body spoken of is separated into its components under the influence of electricity: it is analogous in its sense and sound to *analyze*, which is derived in a similar manner. The term *electrolytical* will be understood at once. Muriatic acid is electrolytical, boracic acid is not.

665. Finally, I require a term to express those bodies which can pass to the *electrodes*, or, as they are usually called, the poles. Substances are frequently spoken of as being *electro-negative*, or *electro-positive*, according as they go under the supposed influence of a direct attraction to the positive or negative pole. But these terms are much too significant for the use to which I should have to put them; for though the meanings are perhaps right, they are only hypothetical, and may be wrong; and then, through a very imperceptible, but still very dangerous, because continual, influence, they do great injury to science, by contracting and limiting the habitual views of those engaged in pursuing it. I propose to distinguish these bodies by calling those *anions** which go to the *anode* of the decomposing body; and those passing to the *cathode*, *cations*†; and when I have occasion to speak of these together, I shall call them *ions*. Thus, the chloride of lead is an *electrolyte*, and when *electrolyzed* evolves the two *ions*, chlorine and lead, the former being an *anion*, and the latter a *cation*.

666. These terms being once well defined, will, I hope, in their use enable me to avoid much periphrasis and ambiguity of expression. I do not mean to press them into service more frequently than will be required, for I am fully aware that names are one thing and science another‡.

667. It will be well understood that I am giving no opinion respecting the nature of the electric current now, beyond what I have done on a former occasion (283. 517.); and that though I speak of the current as proceeding from the parts which are positive to those which are negative (663.), it is merely in accordance with the conventional, though in some degree tacit, agreement entered into by scientific men, that they may have a constant, certain, and definite means of referring to the direction of the forces of that current.

¶ iv. *On some general conditions of Electro-chemical Decomposition.*

669. From the period when electro-chemical decomposition was first effected to the present time, it has been a remark, that those elements which, in the ordinary phenomena of chemical affinity, were the most directly opposed to each other, and combined with the greatest attractive force, were those which were the most readily evolved at the opposite extremities of the decomposing bodies (549.).

* ἀνιον *that which goes up*. (Neuter participle.) † κατιον *that which goes down*.

‡ Since this paper was read, I have changed some of the terms which were first proposed, that I might employ only such as were at the same time simple in their nature, clear in their reference, and free from hypothesis.

670. If this result was evident when water was supposed to be essential to, and was present, in almost every case of such decomposition (472.), it is far more evident now that it has been shown and proved that water is not necessarily concerned in the phenomena (474.), and that other bodies much surpass it in some of the effects supposed to be peculiar to that substance.

671. Water, from its constitution and the nature of its elements, and from its frequent presence in cases of electrolytic action, has hitherto stood foremost in this respect. Though a compound formed by very powerful affinity, it yields up its elements under the influence of a very feeble electric current; and it is doubtful whether a case of electrolyzation can occur, where, being present, it is not resolved into its first principles.

672. The various oxides, chlorides, iodides, and salts (402.), which I have shown are decomposable by the electric current when in the liquid state, under the same general law with water, illustrate in an equally striking manner the activity, in such decompositions, of elements directly and powerfully opposed to each other by their chemical relations.

673. On the other hand, bodies dependent on weak affinities very rarely give way. Take, for instance, glasses: many of those formed of silica, lime, alkali, and oxide of lead, may be considered as little more than solutions of substances one in another*. If bottle-glass be fused, and subjected to the voltaic pile, it does not appear to be at all decomposed (408.). If flint-glass, which contains substances more directly opposed, be operated upon, it suffers some decomposition; and if borate of lead glass, which is a definite chemical compound, be experimented with, it readily yields up its elements (408).

674. But the result which is found to be so striking in the instances quoted is not at all borne out by reference to other cases where a similar consequence might have been expected. It may be said, that my own theory of electro-chemical decomposition would lead to the expectation that all compound bodies should give way under the influence of the electric current with a facility proportionate to the strength of the affinity by which their elements, either proximate or ultimate, are combined. I am not sure that that follows as a consequence of the theory; but if the objection be supposed one presented by facts, I have no doubt it will be removed when we obtain a more intimate acquaintance with, and precise idea of, the nature of chemical affinity and the mode of action of an electric current over it (518. 524.): besides which, it is just as directly opposed to any other theory of electro-chemical decomposition as the one I have propounded; for if it be admitted, as is generally the case, that the more directly bodies are opposed to each other in their attractive forces, the more powerfully do they combine, then the objection applies with equal force to any of the theories of electrolyzation which have been considered, and is an addition to those which I have taken against them.

* Philosophical Transactions, 1830, p. 49.

675. Amongst powerful compounds which are not decomposed, boracic acid stands prominent (408.). Then again, the iodide of sulphur, and the chlorides of sulphur, phosphorus, and carbon, are not decomposable under common circumstances, though their elements are of a nature which would lead to a contrary expectation. Chloride of antimony (402. 690.), the hydro-carbons, acetic acid, ammonia, and many other bodies undecomposable by the voltaic pile, would seem to be formed by an affinity sufficiently strong to indicate that the elements were so far contrasted in their nature as to sanction the expectation that the pile would separate them, especially as in some cases of mere solution (530. 544.), where the affinity must by comparison be very weak, separation takes place*.

676. It must not be forgotten, however, that much of this difficulty, and perhaps the whole, may depend upon the absence of conducting power, which, preventing the transmission of the current, prevents of course the effects due to it. All known compounds being non-conductors when solid, but conductors when liquid, are decomposed, with *perhaps* the single exception at present known of periodide of mercury (679. 691.); and even water itself, which so easily yields up its elements when the current passes, if rendered quite pure, scarcely suffers change, because it then becomes a very bad conductor.

677. If it should hereafter be proved that the want of decomposition in those cases where, from chemical considerations, it might be so strongly expected (669. 674. 672.), is due to the absence or deficiency of conducting power, it would also be proved, at the same time, that decomposition *depends* upon conduction, and not the latter upon the former (413.); and in water this seems to be very nearly decided. On the other hand, the conclusion is almost irresistible, that in electrolytes the power of transmitting the electricity across the substance is dependent upon their capability of suffering decomposition; taking place only whilst they are decomposing, and being proportionate to the quantity of elements separated (821.). I may not, however, stop to discuss this point experimentally at present.

678. When a compound contains such elements as are known to pass towards the opposite extremities of the voltaic pile, still the proportions in which they are present appear to be intimately connected with capability in the compound of suffering or resisting decomposition. Thus, the protochloride of tin readily conducts, and is decomposed (402.), but the perchloride neither conducts nor is decomposed (406.). The protiodide of tin is decomposed when fluid (402.); the periodide is not (405.). The periodide of mercury when fused is not decomposed (691.), even though it does conduct. I was unable to contrast it with the protiodide, the latter being converted into mercury and periodide by heat.

679. These important differences induced me to look more closely to certain binary compounds, with a view of ascertaining whether a *law* regulating the *decomposability*

* With regard to solution, I have met with some reasons for supposing that it will probably disappear as a cause of transference, and intend resuming the consideration at a convenient opportunity.

according to some *relation of the proportionals or equivalents* of the elements, could be discovered. The proto compounds only, amongst those just referred to, were decomposable; and on referring to the substances quoted to illustrate the force and generality of the law of conduction and decomposition which I discovered (402.), it will be found that all the oxides, chlorides, and iodides subject to it, except the chloride of antimony and the periodide of mercury, (to which may now perhaps be added corrosive sublimate,) are also decomposable, whilst many per compounds of the same elements, not subject to the law, were not so (405. 406.).

680. The substances which appeared to form the strongest exceptions to this gegeral result were such bodies as the sulphuric, phosphoric, nitric, arsenic, and other acids.

681. On experimenting with sulphuric acid, I found no reason to believe that it was by itself a conductor of, or decomposable by, electricity, although I had previously been of that opinion (552.). When very strong it is a much worse conductor than if diluted*. If then subjected to the action of a powerful battery, oxygen appears at the *anode*, or positive electrode, although much is absorbed (728.), and hydrogen and sulphur appear at the *cathode*, or negative electrode. Now the hydrogen has with me always been pure, not sulphuretted, and has been deficient in proportion to the sulphur present, so that it is evident that when decomposition occurred water must have been decomposed. I endeavoured to make the experiment with anhydrous sulphuric acid. It appeared to me that in that state, when fused, sulphuric acid was not a conductor, nor decomposed; but I had not enough of the dry acid in my possession to allow me to decide the point satisfactorily. My belief is, that when sulphur appears by the action of the pile on sulphuric acid, it is the result of a secondary action, and that the acid itself is not electrolyzable (757.).

682. Phosphoric acid is, I believe, also in the same condition; but I have found it impossible to decide the point, because of the difficulty of operating on fused anhydrous phosphoric acid. Phosphoric acid which has once obtained water cannot be deprived of it by heat alone. When heated, the hydrated acid volatilizes. Upon subjecting phosphoric acid, fused upon the ring end of a wire (401.), to the action of the voltaic apparatus, it conducted, and was decomposed; but gas, which I believe to be hydrogen, was always evolved at the negative electrode, and the wire was not affected as would have happened had phosphorus been separated. Gas was also evolved at the positive electrode. From all the facts, I conclude it was the water and not the acid which was decomposed.

683. *Arsenic acid.* This substance conducted, and was decomposed; but it contained water, and I was unable at the time to press the investigation so as to ascertain whether a fusible anhydrous arsenic acid could be obtained. It forms, therefore, at present no exception to the general result.

684. Nitrous acid, obtained by distilling nitrate of lead, and keeping it in contact

* De la Rive.

with strong sulphuric acid, was found to conduct and decompose slowly. But on examination there were strong reasons for believing that water was present, and that the decomposition and conduction depended upon it. I endeavoured to prepare a perfectly anhydrous portion, but could not spare the time required to procure an unexceptionable result.

685. Nitric acid is a substance which I believe is not decomposed directly by the electric current. As I want the facts in illustration of the distinction existing between primary and secondary decomposition, I will merely refer to them in this place (752.).

686. That these mineral acids should confer facility of conduction and decomposition on water, is no proof that they are competent to favour and suffer these actions in themselves. Boracic acid does the same thing, though not decomposable. M. De la Rive has pointed out that chlorine has this power also; but being to us an elementary substance, it cannot be due to its capability of suffering decomposition.

687. *Chloride of sulphur* does not conduct, nor is it decomposed. It consists of single proportionals of its elements, but is not on that account an exception to the rule (679.), which does not affirm that *all* compounds of single proportionals of elements are decomposable, but that such as are decomposable are so constituted.

688. *Protochloride of phosphorus* does not conduct nor become decomposed.

689. *Protochloride of carbon* does not conduct nor suffer decomposition. In association with this substance, I submitted the *hydro-chloride of carbon* from olefiant gas and chlorine to the action of the electric current; but it also refused to conduct or yield up its elements.

690. With regard to the exceptions (679.), upon closer examination, some of them disappear. Chloride of antimony (a compound of one proportional of antimony and one and a half of chlorine) of recent preparation was put into a tube (fig. 13.) (789.), and submitted when fused to the action of the current, the positive electrode being of plumbago. No electricity passed, and no appearance of decomposition was visible at first; but when the positive and negative electrodes were brought very near each other in the chloride, then a feeble action occurred and a feeble current passed. The effect altogether was so small (although quite amenable to the law before given), and so unlike the decomposition and conduction occurring in all the other cases, that I attribute it to the presence of a minute quantity of water, (for which this and many other chlorides have strong attractions, producing hydrated chlorides,) or perhaps of a true protochloride consisting of single proportionals (695. 796.).

691. *Periodide of mercury* being examined in the same manner, was found most distinctly to insulate whilst solid, but conduct when fluid, according to the law of *liquido-conduction* (402.); but there was no appearance of decomposition. No iodine appeared at the *anode*, nor mercury or other substance at the *cathode*. The case is, therefore, no exception to the rule, that only compounds of single proportionals are decomposable; but it is an exception, and I think the only one, to the statement, that

all bodies subject to the law of liquido-conduction are decomposable. I incline, however, to believe, that a portion of protiodide of mercury is retained dissolved in the periodide, and that to its slow decomposition the feeble conducting power is due. Periodide would be formed, as a secondary result, at the *anode* ; and the mercury at the *cathode* would also form, as a secondary result, protiodide. Both these bodies would mingle with the fluid mass, and thus no final separation appear, notwithstanding the continued decomposition.

692. When *perchloride of mercury* was subjected to the voltaic current, it did not conduct in the solid state, but it did conduct when fluid. I think, also, that in the latter case it was decomposed; but there are many interfering circumstances which require examination before a positive conclusion can be drawn.

693. When the ordinary protoxide of antimony is subjected to the voltaic current in a fused state, it also is decomposed, although the effect from other causes soon ceases (402. 802.). This oxide consists of one proportional of antimony and one and a half of oxygen, and is therefore an exception to the general law assumed. But in working with this oxide and the chloride, I observed facts which lead me to doubt whether the compounds usually called the protoxide and the protochloride do not often contain other compounds, consisting of single proportions, which are the true proto compounds, and which, in the case of the oxide, might give rise to the decomposition above described.

694. The ordinary sulphuret of antimony is considered as being the compound with the smallest quantity of sulphur, and analogous in its proportions to the ordinary protoxide. But I find that if it be fused with metallic antimony, a new sulphuret is formed, containing much more of the metal than the former, and separating distinctly, when fused, both from the pure metal on the one hand, and the ordinary grey sulphuret on the other. In some rough experiments, the metal thus taken up by the ordinary sulphuret of antimony was equal to half the proportion of that previously in the sulphuret, in which case the new sulphuret would consist of *single* proportionals.

695. When this new sulphuret was dissolved in muriatic acid, although a little antimony separated, yet it appeared to me that a true protochloride, consisting of *single* proportionals, was formed, and from that, by alkalies, &c., a true protoxide, consisting also of *single* proportionals was obtainable. But I could not stop to ascertain this matter strictly by analysis.

696. I believe, however, that there is such an oxide; that it is often present in variable proportions in what is commonly called protoxide, throwing uncertainty upon the results of its analysis, and causing the electrolytic decomposition above described.

697. Upon the whole, it appears probable that all those binary compounds of elementary bodies which are capable of being electrolyzed when fluid, but not whilst solid, according to the law of liquido-conduction (394.), consist of single proportionals of their elementary principles; and it may be because of their departure from this

simplicity of composition, that boracic acid, ammonia, perchlorides, periodides, and many other direct compounds of elements, are indecomposable.

698. With regard to salts and combinations of compound bodies, the same simple relation does not appear to hold good. I could not decide this by bisulphates of the alkalies, for as long as the second proportion of acid remained, water was retained with it. The fused salt, therefore, conducted, and was decomposed; but hydrogen always appeared at the negative electrode.

699. A biphosphate of soda was prepared by heating, and ultimately fusing, the ammonia-phosphate of soda. In this case the fused bisalt conducted, and was decomposed; but a little gas appeared at the negative electrode, and though I believe the salt itself was electrolyzed, I am not quite satisfied that water was entirely absent.

700. Then a biborate of soda was prepared; and this, I think, is an unobjectionable case. The salt, when fused, conducted, and was decomposed, and gas appeared at both electrodes: even when the boracic acid was increased to three proportionals the same effect took place.

701. Hence this class of compound combinations does not seem to be subject to the same simple law as the former class of binary combinations. Whether we may find reason to consider them as mere solutions of the compound of single proportionals in the excess of acid, is a matter which, with some apparent exceptions occurring amongst the sulphurets, must be left for decision by future examination.

702. In any investigation of these points, great care must be taken to exclude water; for if present, secondary effects are so frequently produced as often seemingly to indicate an electro-decomposition of substances, when no true result of the kind has occurred (742. &c.).

703. It is evident that all the cases in which decomposition *does not occur may* depend upon the want of conduction (677. 413.); but that does not at all lessen the interest excited by seeing the great difference of effect due to a change, not in the nature of the elements, but merely in their proportions, especially in any attempt which may be made to elucidate and expound the beautiful theory put forth by Sir HUMPHRY DAVY*, and illustrated by BERZELIUS and other eminent philosophers, that ordinary chemical affinity is a mere result of the electrical attractions of the particles of matter.

¶ v. *On a new Measurer of Volta-electricity.*

704. I have already said, when engaged in reducing common and voltaic electricity to one standard of measurement (377.), and again when introducing my theory of electro-chemical decomposition (504. 505. 510.), that the chemical decomposing action of a current *is constant for a constant quantity of electricity,* notwithstanding the greatest variations in its sources, in its intensity, in the size of the *electrodes* used, in the nature of the conductors (or non-conductors (307.)) through which it is

* Philosophical Transactions, 1807, pp. 32, 39; also 1826, pp. 387, 389.

passed, or in other circumstances. The conclusive proofs of the truth of these statements shall be given almost immediately (783. &c.).

705. I endeavoured upon this law to construct an instrument which should measure out the electricity passing through it, and which, being interposed in the course of the current used in any particular experiment, should serve at pleasure, either as a *comparative standard* of effect, or as a *positive measurer* of this subtile agent.

706. There is no substance better fitted, under ordinary circumstances, to be the indicating body in such an instrument than water; for it is decomposed with facility when rendered a better conductor by the addition of acids or salts; its elements may in numerous cases be obtained and collected without any embarrassment from secon dary action, and, being gaseous, they are in the best physical condition for separation and measurement. Water, therefore, acidulated by sulphuric acid, is the substance I shall generally refer to, although it may become expedient in peculiar cases or forms of experiment to use other bodies (843.).

707. The first precaution needful in the construction of the instrument was to avoid the recombination of the evolved gases, an effect which the positive electrode has been found so capable of producing (571.). For this purpose various forms of decomposing apparatus were used. The first consisted of straight tubes, each containing a plate and wire of platina soldered together by gold, and fixed hermetically in the glass at the closed extremity of the tube (Plate I. fig. 5.). The tubes were about eight inches long, 0·7 of an inch in diameter, and graduated. The platina plates were about an inch long, as wide as the tubes would permit, and adjusted as near to the mouths of the tubes as was consistent with the safe collection of the gases evolved. In certain cases, where it was required to evolve the elements upon as small a surface as possible, the metallic extremity, instead of being a plate, consisted of the wire bent into the form of a ring (fig. 6.). When these tubes were used as measurers, they were filled with the dilute sulphuric acid, and inverted in a basin of the same liquid (fig. 7.), being placed in an inclined position, with their mouths near to each other, that as little decomposing matter should intervene as possible; and also, in such a direction that the platina plates should be in vertical planes (720.).

708. Another form of apparatus was that delineated (fig. 8.). The tube is bent in the middle; one end is closed; in that end is fixed a wire and plate, *a*, proceeding so far downwards, that, when in the position figured, it shall be as near to the angle as possible, consistently with the collection, at the closed extremity of the tube, of all the gas evolved against it. The plane of this plate is also perpendicular (720.). The other metallic termination, *b*, is introduced at the time decomposition is to be effected, being brought as near the angle as possible, without causing any gas to pass from it towards the closed end of the instrument. The gas evolved against it is allowed to escape.

709. The third form of apparatus contains both electrodes in the same tube; the transmission, therefore, of the electricity, and the consequent decomposition, is far

more rapid than in the separate tubes. The resulting gas is the sum of the portions evolved at the two electrodes, and the instrument is better adapted than either of the former as a measurer of the quantity of voltaic electricity transmitted in ordinary cases. It consists of a straight tube (fig. 9.) closed at the upper extremity, and graduated, through the sides of which pass the platina wires (being fused into the glass), which are connected with two plates within. The tube is fitted by grinding into one mouth of a double-necked bottle. If the latter be one half or two thirds full of the dilute sulphuric acid, it will, upon inclination of the whole, flow into the tube and fill it. When an electric current is passed through the instrument, the gases evolved against the plates collect in the upper portion of the tube, and are not subject to the recombining power of the platina.

710. Another form of the instrument is given at fig. 10.

711. A fifth form is delineated (fig. 11.). This I have found exceedingly useful in experiments continued in succession for days together, and where large quantities of indicating gas were to be collected. It is fixed on a weighted foot, and has the form of a small retort containing the two electrodes: the neck is narrow, and sufficiently long to deliver gas issuing from it into a jar placed in a small pneumatic trough. The electrode chamber, sealed hermetically at the part held in the stand, is five inches in length, and 0·6 of an inch in diameter; the neck about nine inches in length, and 0·4 of an inch in diameter internally. The figure will fully indicate the construction.

712. It can hardly be requisite to remark, that in the arrangement of any of these forms of apparatus, they, and the wires connecting them with the substance, which is collaterally subjected to the action of the same electric current, should be so far insulated as to ensure a certainty that all the electricity which passes through the one shall also be transmitted through the other.

713. Next to the precaution of collecting the gases, if mingled, out of contact with the platinum, was the necessity of testing the law of a *definite electrolytic* action, upon water at least, under all varieties of condition; that, with a conviction of its certainty, might also be obtained a knowledge of those interfering circumstances which would require to be practically guarded against.

714. The first point investigated was the influence or indifference of extensive variations in the size of the electrodes, for which purpose instruments like those last described (709. 710. 711.) were used. One of these had plates 0·7 of an inch wide, and nearly four inches long; another had plates only 0·5 of an inch wide, and 0·8 of an inch long; a third had wires 0·02 of an inch in diameter, and three inches long; and a fourth similar wires only half an inch in length. Yet when these were filled with dilute sulphuric acid, and, being placed in succession, had one common current of electricity passed through them, very nearly the same quantity of gas was evolved in all. The difference was sometimes in favour of one, and sometimes on the side of another; but the general result was that the largest quantity of gases was evolved upon the smaller surface of the wires.

715. Experiments of a similar kind were made with the single-plate, straight tubes (707.), and also with the curved tubes (708.), with similar consequences; and when these, with the former tubes, were arranged together in various ways, the result, as to the equality of action of large and small metallic surfaces when delivering and receiving the same current of electricity, was constantly the same. As an illustration, the following numbers are given. An instrument with two wires evolved 74·3 volumes of mixed gases; another with plates 73·25 volumes; whilst the sum of the oxygen and hydrogen in two separate tubes amounted to 73·65 volumes. In an other experiment the volumes were 55·3, 55·3, and 54·4.

716. But it was observed in these experiments, that in single-plate tubes (707.) more hydrogen was evolved at the negative electrode than was proportionate to the oxygen at the positive electrode; and generally, also, more than was proportionate to the oxygen and hydrogen in a double-plate tube. Upon more minutely examining these effects, I was led to refer them, and also the differences between wires and plates (714.), to the solubility of the gases evolved, especially at the positive electrode.

717. When the positive and negative electrodes are equal in surface, the bubbles which rise from them in dilute sulphuric acid are always different in character. Those from the positive plate are exceedingly small, and separate instantly from every part of the surface of the metal, in consequence of its perfect cleanliness (633.); whilst in the liquid they give it a hazy appearance, from their number and minuteness; are easily carried down by currents; and therefore not only present far greater surface of contact with the liquid than larger bubbles would do, but are retained a much longer time in mixture with it. But the bubbles at the negative surface, though they constitute twice the volume of the gas at the positive electrode, are nevertheless very inferior in number. They do not rise so universally from every part of the surface, but seem to be evolved at different points; and though so much larger, they appear to cling to the metal, separating with difficulty from it, and when separated, instantly rising to the top of the liquid. If, therefore, oxygen and hydrogen had equal solubility in, or powers of combining with, water under similar circumstances, still under the present conditions the oxygen would be far the most liable to solution; but when to these is added its well known power of forming a compound with water, it is no longer surprising that such a compound should be produced in small quantities at the positive electrode; and indeed the bleaching power which some philosophers have observed in a solution at this electrode, when chlorine and similar bodies have been carefully excluded, is probably due to the formation there, in this manner, of oxy-water.

718. That more gas was collected from the wires than from the plates, I attribute to the circumstance, that as equal quantities were evolved in equal times, the bubbles at the wires having been more rapidly produced, in relation to any part of the surface, must have been much larger; have been therefore in contact with the fluid by a much

smaller surface, and for a much shorter time than those at the plates; hence less solution and a greater collection.

719. There was also another effect produced, especially by the use of large electrodes, which was both a consequence and a proof of the solution of part of the gas evolved there. The collected gas, when examined, was found to contain small portions of nitrogen. This I attribute to the presence of air dissolved in the acid used for decomposition. It is a well-known fact, that when bubbles of a gas but slightly soluble in water or solutions pass through them, the portion of this gas which is dissolved displaces a portion of that previously in union with the liquid: and so, in the decompositions under consideration, as the oxygen dissolves, it displaces a part of the air, or at least of the nitrogen, previously united to the acid; and this proceeds *most extensively* with large plates, because the gas evolved at them is in the most favourable condition for solution.

720. With the intention of avoiding this solubility of the gases as much as possible, I arranged the decomposing plates in a vertical position (707. 708.), that the bubbles might quickly escape upwards, and that the downward currents in the fluid should not meet ascending currents of gas. This precaution I found to assist greatly in producing constant results, and especially in experiments to be hereafter referred to, in which other liquids than dilute sulphuric acid, as for instance solution of potash, were used.

721. The irregularities in the indications of the measurer proposed, arising from the solubility just referred to, are but small, and may be very nearly corrected by comparing the results of two or three experiments. They may also be almost entirely avoided by selecting that solution which is found to favour them in the least degree (728.); and still further by collecting the hydrogen only, and using that as the indicating gas; for being much less soluble than oxygen, being evolved with twice the rapidity and in larger bubbles (717.), it can be collected more perfectly and in greater purity.

722. From the foregoing and many other experiments, it results that *variation in the size of the electrodes causes no variation in the chemical action of a given quantity of electricity upon water.*

723. The next point in regard to which the principle of constant electro-chemical action was tested, was *variation of intensity.* In the first place, the preceding experiments were repeated, using batteries of an *equal* number of plates, *strongly* and *weakly* charged; but the results were alike. They were then repeated, using batteries sometimes containing forty, and at other times only five pairs of plates; but the results were still the same. *Variations therefore in the intensity,* caused by difference in the strength of charge, or in the number of alternations used, *produced no difference as to the equal action of large and small electrodes.*

724. Still these results did not prove that variation in the intensity of the current was not accompanied by a corresponding variation in the electro-chemical effects,

since the actions at *all* the surfaces might have increased or diminished together. The deficiency in the evidence is, however, completely supplied by the former experiments on different-sized electrodes; for with variation in the size of these, a variation in the intensity must have occurred. The intensity of an electric current traversing conductors alike in their nature, quality, and length, is probably as the quantity of electricity passing through a given sectional area perpendicular to the current, divided by the time (360. *note*); and therefore when large plates were contrasted with wires separated by an equal length of the same decomposing conductor (714.), whilst one current of electricity passed through both arrangements, that electricity must have been in a very different state, as to *tension*, between the plates and between the wires; yet the chemical results were the same.

725. The difference in intensity, under the circumstances described, may be easily shown practically, by arranging two decomposing apparatus as in fig. 12, where the same fluid is subjected to the decomposing power of the same current of electricity, passing in the vessel A. between large platina plates, and in the vessel B. between small wires. If a third decomposing apparatus, such as that delineated fig. 11. (711.), be connected with the wires at *a b*, fig. 12, it will serve sufficiently well, by the degree of decomposition occurring in it, to indicate the relative state of the two plates as to intensity; and if it then be applied in the same way, as a test of the state of the wires at *a′ b′*, it will, by the increase of decomposition within, show how much greater the intensity is there than at the former points. The connexions of P and N with the voltaic battery are of course to be continued during the whole time.

726. A third form of experiment in which difference of intensity was obtained, for the purpose of testing the principle of equal chemical action, was to arrange three volta-electrometers, so that after the electric current had passed through one, it should divide into two parts, which, after traversing each one of the remaining instruments, should reunite. The sum of the decomposition in the two latter vessels was always equal to the decomposition in the former vessel. But the *intensity* of the divided current could not be the same as that it had in its original state; and therefore *variation of intensity has no influence on the results if the quantity of electricity remain the same.* The experiment, in fact, resolves itself simply into an increase in the size of the electrodes (725.).

727. The *third point*, in respect to which the principle of equal electro-chemical action on water was tested, was *variation of the strength of the solution used.* In order to render the water a conductor, sulphuric acid had been added to it (707.); and it did not seem unlikely that this substance, with many others, might render the water more subject to decomposition, the electricity remaining the same in quantity. But such did not prove to be the case. Diluted sulphuric acid, of different strengths, was introduced into different decomposing apparatus, and submitted simultaneously to the action of the same electric current (714.). Slight differences occurred, as before, sometimes in one direction, sometimes in another; but the final result was, that

exactly the same quantity of water was decomposed in all the solutions by the same quantity of electricity, though the sulphuric acid in some was seventyfold what it was in others. The strengths used were of specific gravity 1·495, and downwards.

728. When an acid having a specific gravity of about 1·336 was employed, the results were most uniform, and the oxygen and hydrogen (716.) most constantly in the right proportion to each other. Such an acid gave more gas than one much weaker acted upon by the same current, apparently because it had less solvent power. If the acid were very strong, then a remarkable disappearance of oxygen took place; thus, one made by mixing two measures of strong oil of vitriol with one of water, gave forty-two volumes of hydrogen, but only twelve of oxygen. The hydrogen was very nearly the same with that evolved from acid of the specific gravity 1·232. I have not yet had time to examine minutely the circumstances attending the disappearance of the oxygen in this case, but imagine it is due to the formation of oxywater, which Thénard has shown is favoured by the presence of acid.

729. Although not necessary for the practical use of the instrument I am describing, yet as connected with the important point of constant electro-chemical action upon water, I now investigated the effects produced by an electric current passing through aqueous solutions of acids, salts, and compounds, exceedingly different from each other in their nature, and found them to yield astonishingly uniform results. But many of them which are connected with a secondary action will be more usefully described hereafter (778.).

730. When solutions of caustic potassa or soda, or sulphate of magnesia, or sulphate of soda, were acted upon by the electric current, just as much oxygen and hydrogen was evolved from them as from the diluted sulphuric acid, with which they were compared. When a solution of ammonia, rendered a better conductor by sulphate of ammonia (554.), or a solution of subcarbonate of potassa was experimented with, the *hydrogen* evolved was in the same quantity as that set free from the diluted sulphuric acid with which they were compared. Hence *changes in the nature of the solution do not alter the constancy of electrolytic action upon water.*

731. I have already said, respecting large and small electrodes, that change of order caused no change in the general effect (715.). The same was the case with different solutions, or with different intensities; and however the circumstances of an experiment might be varied, the results came forth exceedingly consistent, and proved that the electro-chemical action was still the same.

732. I consider the foregoing investigation as sufficient to prove the very extraordinary and important principle with respect to water, *that when subjected to the influence of the electric current, a quantity of it is decomposed exactly proportionate to the quantity of electricity which has passed,* notwithstanding the thousand variations in the conditions and circumstances under which it may at the time be placed; and further, that when the interference of certain secondary effects (742. &c.), together with the solution or recombination of the gas and the evolution of air, are guarded against,

the products of the decomposition may be collected with such accuracy, as to afford a very excellent and valuable measurer of the electricity concerned in their evolution.

733. The forms of instrument which I have given, figg. 9, 10, 11. (709. 710. 711.), are probably those which will be found most useful, as they indicate the quantity of electricity by the largest volume of gases, and cause the least obstruction to the passage of the current. The fluid which my present experience leads me to prefer, is a solution of sulphuric acid of specific gravity about 1·336, or from that to specific gravity 1·25; but it is very essential that there should be no organic substance, nor any vegetable acid, nor other body, which, by being liable to the action of the oxygen or hydrogen evolved at the electrodes (773. &c.), shall diminish their quantity, or add other gases to them.

734. In many cases when the instrument is used as a *comparative standard*, or even as *a measurer*, it may be desirable to collect the hydrogen only, as being less liable to absorption or disappearance in other ways than the oxygen; whilst at the same time its volume is so large, as to render it a good and sensible indicator. In such cases the first and second form of apparatus have been used, figg. 7, 8. (707. 708.). The indications obtained were very constant, the variations being much smaller than in those forms of apparatus collecting both gases; and they can also be procured when solutions are used in comparative experiments, which, yielding no oxygen or only secondary results of its action, can give no indications if the educts at both electrodes be collected. Such is the case when solutions of ammonia, muriatic acid, chlorides, iodides, acetates, or other vegetable salts, &c., are employed.

735. In a few cases, as where solutions of metallic salts liable to reduction at the negative electrode are acted upon, the oxygen may be advantageously used as the measuring substance. This is the case, for instance, with sulphate of copper.

736. There are therefore two general forms of the instrument which I submit as a measurer of electricity. One, in which both the gases of the water decomposed are collected (709. 710. 711.); and the other, in which a single gas, as the hydrogen only, is used (707. 708.). When referred to as a *comparative instrument*, (a use I shall now make of it very extensively,) it will not often require particular precaution in the observation; but when used as an *absolute measurer*, it will be needful that the barometric pressure and the temperature be taken into account, and that the graduation of the instruments should be to one scale; the hundredths and smaller divisions of a cubical inch are quite fit for this purpose, and the hundredth may be very conveniently taken as indicating a DEGREE of electricity.

737. It can scarcely be needful to point out further than has been done how this instrument is to be used. It is to be introduced into the course of the electric current, the action of which is to be exerted anywhere else, and if 60° or 70° of electricity are to be measured out, either in one or several portions, the current, whether strong or weak, is to be continued until the gas in the tube occupies that number of divisions or hundredths of a cubical inch. Or if a quantity competent to produce a certain

effect is to be measured, the effect is to be obtained, and then the indication read off. In exact experiments it is necessary to correct the volume of gas for changes in temperature and pressure, and especially for moisture*. For the latter object the volta-electrometer (fig. 11.) is most accurate, as its gas can be measured over water, whilst the others retain it over acid or saline solutions.

738. I have not hesitated to apply the term *degree*, in analogy with the use made of it with respect to another most important imponderable agent, namely, heat; and as the definite expansion of air, water, mercury, &c., is there made use of to measure heat, so the equally definite evolution of gases is here turned to a similar use for electricity.

739. The instrument offers the only *actual measurer* of voltaic electricity which we at present possess. For without being at all affected by variations in time or intensity, or alterations in the current itself, of any kind, or from any cause, or even of intermissions of action, it takes note with accuracy of the quantity of electricity which has passed through it, and reveals that quantity by inspection; I have therefore named it a VOLTA-ELECTROMETER.

740. Another mode of measuring volta-electricity may be adopted with advantage in many cases, dependent on the quantities of metals or other substances evolved either as primary or as secondary results; but I refrain from enlarging on this use of the products, until the principles on which their constancy depends have been fully established (791. 843.).

741. By the aid of this instrument I have been able to establish the definite character of electro-chemical action in its most general sense; and I am persuaded it will become of the utmost use in the extensions of the science which these views afford. I do not pretend to have made its detail perfect, but to have demonstrated the truth of the principle, and the utility of the application.

¶ vi. *On the primary or secondary character of the bodies evolved at the Electrodes.*

742. Before the *volta-electrometer* could be employed in determining, as a *general law*, the constancy of electro-decomposition, it became necessary to examine a distinction, already recognised among scientific men, relative to the products of that action, namely, their primitive or secondary character; and, if possible, by some general rule or principle, to decide when they were of the one or the other kind. It will appear hereafter that great mistakes respecting electro-chemical action and its consequences, have arisen from confounding these two classes of results together.

743. When a substance under decomposition yields at the electrodes those bodies uncombined and unaltered which the electric current has separated, then they may be considered as primary results, even though themselves compounds. Thus the oxygen and hydrogen from water are primary results; and so also are the acid and alkali (themselves compound bodies) evolved from sulphate of soda. But when the sub-

* For a simple table of correction for moisture, I may take the liberty of referring to my Chemical Manipulation, edition of 1830, p. 376.

stances separated by the current are changed at the electrodes before their appearance, then they give rise to secondary results, although in many cases the bodies evolved are elementary.

744. These secondary results occur in two ways, being sometimes due to the mutual action of the evolving substance and the matter of the electrode, and sometimes to its action upon the substances contained in the decomposing conductor itself. Thus, when carbon is made the positive electrode in dilute sulphuric acid, carbonic oxide and carbonic acid appear there instead of oxygen; for the latter, acting upon the matter of the electrode, produces these secondary results. Or if the positive electrode, in a solution of nitrate or acetate of lead, be platina, then peroxide of lead appears there, equally a secondary result with the former, but now depending upon an action of the oxygen on a substance in the solution. Again, when ammonia is decomposed by platina electrodes, nitrogen appears at the *anode** ; but though an *elementary* body, it is a *secondary* result in this case, being derived from the chemical action of the oxygen electrically evolved there, upon the ammonia in the surrounding solution (554.). In the same manner when aqueous solutions of metallic salts are decomposed by the current, the metals evolved at the *cathode*, though elements, are *always* secondary results, and not immediate consequences of the decomposing power of the electric current.

745. Many of these secondary results are extremely valuable; for instance, all the interesting compounds which M. Becquerel has obtained by feeble electric currents are of this nature; but they are essentially chemical, and must, in the theory of electrolytic action, be carefully distinguished from those which are directly due to the action of the electric current.

746. The nature of the substances evolved will often lead to a correct judgement of their primary or secondary character, but is not sufficient alone to establish that point. Thus, nitrogen is said to be attracted sometimes by the positive and sometimes by the negative electrode, according to the bodies with which it may be combined (554. 555.), and it is on such occasions evidently viewed as a primary result†; but I think I shall show, that, when it appears at the positive electrode, or rather at the *anode*, it is a secondary result (748.). Thus, also, Sir Humphry Davy ‡, and with him the great body of chemical philosophers, (including myself,) have given the appearance of copper, lead, tin, silver, gold, &c., at the negative electrode, when their aqueous solutions were acted upon by the voltaic current, as proofs that the metals, as a class, were attracted to that surface; thus assuming the metal in each case to be a primary result. These however, I expect to prove, are all secondary results; the mere consequence of chemical action, and no proofs of the attraction or the law announced §.

* Annales de Chimie, 1804, tom. li. p. 167.

† Ibid. tom. li. p. 172.

‡ Elements of Chemical Philosophy, pp. 144. 161.

§ It is remarkable that up to 1804 it was the received opinion that the metals were reduced by the nascent hydrogen. At that date the general opinion was reversed by Hisinger and Berzelius (Annales de Chimie,

747. But when we take to our assistance the law of *constant electro-chemical action* already proved with regard to water (732.), and which I hope to extend satisfactorily to all bodies (821.), and consider the *quantities* as well as the *nature* of the substances set free, a generally accurate judgement of the primary or secondary character of the results may be formed: and this important point, so essential to the theory of electro-decomposition, since it decides what are the particles directly under the influence of the current, (distinguishing them from such as are not affected,) and what are the results to be expected, may be established with such degree of certainty as to remove innumerable ambiguities and doubtful considerations from this branch of the science.

748. Let us apply these principles to the case of ammonia, and the supposed determination of nitrogen to one or the other *electrode* (554. 555.). A pure strong solution of ammonia is as bad a conductor, and therefore as little liable to electro-decomposition, as pure water; but when sulphate of ammonia is dissolved in it, the whole becomes a conductor; nitrogen *almost* and occasionally *quite* pure is evolved at the *anode*, and hydrogen at the *cathode*; the ratio of the volume of the former to that of the latter varying, but being as 1 to about 3 or 4. This result would seem at first to imply that the electric current had decomposed ammonia, and that the nitrogen had been determined towards the positive electrode. But when the electricity used was measured out by the volta-electrometer (707. 736.), it was found that the hydrogen obtained was exactly in the proportion which would have been supplied by decomposed water, whilst the nitrogen had no certain or constant relation whatever. When, upon multiplying experiments, it was found that, by using a stronger or weaker solution, or a more or less powerful battery, the gas evolved at the *anode* was a mixture of oxygen and nitrogen, varying both in proportion and absolute quantity, whilst the hydrogen at the *cathode* remained constant, no doubt could be entertained that the nitrogen at the *anode* was a secondary result, depending upon the chemical action of the nascent oxygen, determined to that surface by the electric current, upon the ammonia in solution. It was the water, therefore, which was electrolyzed, not the ammonia. Further, the experiment gives no real indication of the tendency of the element nitrogen to either one electrode or the other; nor do I know of any experiment with nitric acid, or other compounds of nitrogen, which shows the tendency of this element, under the influence of the electric current, to pass in either direction along its course.

749. As another illustration of secondary results, the effects on a solution of acetate of potassa may be quoted. When a very strong solution was used, more gas was evolved at the *anode* than at the *cathode*, in the proportion of 4 to 3 nearly: that from the *anode* was a mixture of carbonic oxide and carbonic acid; that from the *cathode* pure hydrogen. When a much weaker solution was used, less gas was evolved at the *anode* than at the *cathode*; and it now contained carburetted hydrogen, as well as carbonic oxide and car-

1804, tom. li. p. 174.), who stated that the metals were evolved directly by the electricity: in which opinion it appears, from that time, DAVY coincided (Philosophical Transactions, 1826, p. 388.).

bonic acid. This result of carburetted hydrogen at the positive electrode has a very anomalous appearance, if considered as an immediate consequence of the decomposing power of the current. It, however, as well as the carbonic oxide and acid, is only a *secondary result*; for it is the water alone which suffers electro-decomposition, and it is the oxygen eliminated at the *anode* which, reacting on the acetic acid, in the midst of which it is evolved, produces those substances that finally appear there. This is fully proved by experiments with the volta-electrometer (707.); for then the hydrogen evolved from the acetate at the *cathode* is always found to be definite, being exactly proportionate to the electricity which has passed through the solution, and, in quantity, the same as the hydrogen evolved in the volta-electrometer itself. The appearance of the carbon in combination with the hydrogen at the positive electrode, and its non-appearance at the negative electrode, are in curious contrast with the results which might have been expected from the law usually accepted respecting the final places of the elements.

750. If the salt in solution be an acetate of lead, then the results at both electrodes are secondary, and cannot be used to estimate or express the amount of electro-chemical action, except by a circuitous process (843.). In place of oxygen, or even the gases already described (749.), peroxide of lead now appears at the positive, and lead itself at the negative electrode. When other metallic solutions are used, containing, for instance, peroxides, as that of copper, combined with this or any other decomposable acid, still more complicated results will be obtained; which, viewed as direct results of the electro-chemical action, will, in their proportions, present nothing but confusion, but will appear perfectly harmonious and simple if they be considered as secondary results, and will accord in their proportions with the oxygen and hydrogen evolved from water by the action of a definite quantity of electricity.

751. I have experimented upon many bodies, with a view to determine whether the results were primary or secondary. I have been surprised to find how many of them, in ordinary cases, are of the latter class, and how frequently water is the only body electrolyzed in instances where other substances have been supposed to give way. Some of these results I will give in as few words as possible.

752. *Nitric acid.*—When very strong, it conducted well, and yielded oxygen at the positive electrode. No gas appeared at the negative electrode; but nitrous acid, and apparently nitric oxide, were formed there, which, dissolving, rendered the acid yellow or red, and at last even effervescent, from the spontaneous separation of nitric oxide. Upon diluting the acid with its bulk or more of water, gas appeared at the negative electrode. Its quantity could be varied by variations, either in the strength of the acid or of the voltaic current: for that acid from which no gas separated at the *cathode*, with a weak voltaic battery, did evolve gas there with a stronger; and that battery which evolved no gas there, with a strong acid, did cause its evolution with an acid more dilute. The gas at the *anode* was always oxygen; that at the *cathode* hydrogen. When the quantity of products was examined by the volta-electro-

meter (707.), the oxygen, whether from strong or weak acid, proved to be in the same proportion as from water. When the acid was diluted to specific gravity 1·24, or less, the hydrogen also proved to be the same in quantity as from water. Hence I conclude that the nitric acid does not undergo electro-chemical decomposition, but the water only; that the oxygen at the *anode* is always a primary result, but that the products at the *cathode* are often secondary, and due to the reaction of the hydrogen upon the nitric acid.

753. *Nitre.*—A solution of this salt yields very variable results, according as one or other form of tube is used, or as the electrodes are large or small. Sometimes the whole of the hydrogen of the water decomposed may be obtained at the negative electrode; at other times, only a part of it, because of the ready formation of secondary results. The solution is a very excellent conductor of electricity.

754. *Nitrate of ammonia*, in aqueous solution, gives rise to secondary results very varied and uncertain in their proportions.

755. *Sulphurous acid.*—Pure liquid sulphurous acid does not conduct nor suffer decomposition by the voltaic current *, but, when dissolved in water, the solution acquires conducting power, and is decomposed, yielding oxygen at the *anode*, and hydrogen and sulphur at the *cathode*.

756. A solution containing sulphuric acid in addition, was a better conductor. It gave very little gas at either electrode: that at the *anode* was oxygen, that at the *cathode* pure hydrogen. From the *cathode* also rose a white turbid stream, consisting of diffused sulphur, which soon rendered the whole solution milky. The volumes of gases were in no regular proportion to the quantities evolved from water in the volta-electrometer. I conclude that the sulphurous acid was not at all affected by the electric current in any of these cases, and that the water present was the only body electro-chemically decomposed; that, at the *anode*, the oxygen from the water converted the sulphurous acid into sulphuric acid, and, at the *cathode*, the hydrogen electrically evolved decomposed the sulphurous acid, combining with its oxygen, and setting its sulphur free. I conclude that the sulphur at the negative electrode was only a secondary result; and, in fact, no part of it was found combined with the small portion of hydrogen which escaped when weak solutions of sulphurous acid were used.

757. *Sulphuric acid.*—I have already given my reasons for concluding that sulphuric acid is not electrolyzable, i. e. not decomposable directly by the electric current, but occasionally suffering by a secondary action at the *cathode* from the hydrogen evolved there (681.). In the year 1800, Davy considered the sulphur from sulphuric acid as the result of the action of the nascent hydrogen †. In 1804, Hisinger and Berzelius stated that it was the direct result of the action of the voltaic pile ‡; an opinion which from that time Davy seems to have adopted, and which has since been

* See also De la Rive, Bibliothèque Universelle, tom. xl. p. 205; or Quarterly Journal of Science, vol. xxvii. p. 407.

† Nicholson's Quarterly Journal, vol. iv. pp. 280, 281.

‡ Annales de Chimie, 1804, tom. li. p. 173.

commonly received by all. The change of my own opinion requires that I should correct what I have already said of the decomposition of sulphuric acid in a former series of these Researches (552.): I do not now think that the appearance of the sulphur at the negative electrode is an immediate consequence of electrolytic action.

758. *Muriatic acid.*—A strong solution gave hydrogen at the negative electrode, and chlorine only at the positive electrode; of the latter, a part acted on the platina and a part was dissolved. A minute bubble of gas remained; it was not oxygen, but probably air previously held in solution.

759. It was an important matter to determine whether the chlorine was a primary result, or only a secondary product, due to the action of the oxygen evolved from water at the *anode* upon the muriatic acid; i. e. whether the muriatic acid was electrolyzable, and if so, whether the decomposition was *definite*.

760. The muriatic acid was gradually diluted. One part with six of water gave only chlorine at the *anode*. One part with eight of water gave only chlorine; with nine of water, a little oxygen appeared with the chlorine: but the occurrence or non-occurrence of oxygen at these strengths depended, in part, on the strength of the voltaic battery used. With fifteen parts of water, a little oxygen, with much chlorine, was evolved at the *anode*. As the solution was now becoming a bad conductor of electricity, sulphuric acid was added to it: this caused more ready decomposition, but did not sensibly alter the proportion of chlorine and oxygen.

761. The muriatic acid was now diluted with 100 times its volume of dilute sulphuric acid. It still gave a large proportion of chlorine at the *anode*, mingled with oxygen; and the result was the same, whether a voltaic battery of 40 pairs of plates or one containing only 5 pairs were used. With acid of this strength, the oxygen evolved at the *anode* was to the hydrogen at the *cathode*, in volume, as 17 is to 64; and therefore the chlorine would have been 30 volumes, had it not been dissolved by the fluid.

762. Next, with respect to the quantity of elements evolved. On using the volta-electrometer, it was found that, whether the strongest or the weakest muriatic acid were used, whether chlorine alone or chlorine mingled with oxygen appeared at the *anode*, still the hydrogen evolved at the *cathode* was a constant quantity, i. e. exactly the *same* as the hydrogen which the *same quantity of electricity* could evolve from water.

763. This constancy does not decide whether the muriatic acid is electrolyzed or not, although it proves that if so, it must be in definite proportions to the quantity of electricity used. Other considerations may, however, be allowed to decide the point. The analogy between chlorine and oxygen, in their relations to hydrogen, is so strong, as to lead almost to the certainty, that, when combined with that element, they would perform similar parts in the process of electro-decomposition. They both unite with it in single proportional or equivalent quantities; and, the number of proportionals appearing to have an intimate and important relation to the decomposability of a

body (697.), those in muriatic acid, as well as in water, are the most favourable, or those, perhaps even necessary, to decomposition. In other binary compounds of chlorine also, where nothing equivocal depending on the simultaneous presence of it and oxygen is involved, the chlorine is directly eliminated at the *anode* by the electric current. Such is the case with the chloride of lead (395.), which may be justly compared with protoxide of lead (402.), and stands in the same relation to it as muriatic acid to water. The chlorides of potassium, sodium, barium, &c., are in the same relation to the protoxides of the same metals, and present the same results under the influence of the electric current (402.).

764. From all the experiments, combined with these considerations, I conclude that muriatic acid is decomposed by the direct influence of the electric current, and that the quantities evolved are, and therefore the chemical action is, *definite for a definite quantity of electricity*. For though I have not collected and measured the chlorine, in its separate state, at the *anode*, there can exist no doubt as to its being proportional to the hydrogen at the *cathode*; and the results are therefore sufficient to establish the general law of *constant electro-chemical action* in the case of muriatic acid.

765. In the dilute acid (761.), I conclude that a part of the water is electro-chemically decomposed, giving origin to the oxygen, which appears mingled with the chlorine at the *anode*. The oxygen *may* be viewed as a secondary result; but I incline to believe that it is not so: for, if it were, it might be expected in largest proportion from the stronger acid, whereas the reverse is the fact. This consideration, with others, also leads me to conclude that muriatic acid is more easily decomposed by the electric current than water; since, even when diluted with eight or nine times its quantity of the latter fluid, it alone gives way, the water remaining unaffected.

766. *Chlorides.*—On using solutions of chlorides in water,—for instance, the chlorides of sodium or calcium,—there was evolution of chlorine only at the positive electrode, and of hydrogen, with the oxide of the base, as soda or lime, at the negative electrode. The process of decomposition may be viewed as proceeding in two or three ways, all terminating in the same results. Perhaps the simplest is to consider the chloride as the substance electrolyzed, its chlorine being determined to and evolved at the *anode*, and its metal passing to the *cathode*, where, finding no more chlorine, it acts upon the water, producing hydrogen and an oxide as secondary results. As the discussion would detain me from more important matter, and is not of immediate consequence, I shall defer it for the present. It is, however, of *great consequence* to state, that, on using the volta-electrometer, the hydrogen in both cases was definite; and if the results do not prove the definite decomposition of chlorides, (which shall be proved elsewhere,—789. 794. 814.,) they are not in the slightest degree opposed to such a conclusion, and *do* support the *general law*.

767. *Hydriodic acid.*—A solution of hydriodic acid was affected exactly in the same manner as muriatic acid. When strong, hydrogen was evolved at the negative electrode, in definite proportion to the quantity of electricity which had passed, i. e. in

the same proportion as was evolved by the same current from water; and iodine with out any oxygen was evolved at the positive electrode. But when diluted, small quantities of oxygen appeared with the iodine at the *anode*, the proportion of hydrogen at the *cathode* remaining undisturbed.

768. I believe the decomposition of the hydriodic acid in this case to be direct, for the reasons already given respecting muriatic acid (763. 764.).

769. *Iodides.*—A solution of iodide of potassium being subjected to the voltaic current, iodine appeared at the positive electrode (without any oxygen), and hydrogen with free alkali at the negative electrode. The same observations as to the mode of decomposition are applicable here as were made in relation to the chlorides when in solution (766.).

770. *Hydro-fluoric acid and fluorides.*—Solution of hydro-fluoric acid did not appear to be decomposed under the influence of the electric current: it was the water which gave way apparently. The fused fluorides were electrolyzed (417.); but having during these actions obtained *fluorine* in the separate state, I think it better to refer to a future series of these Researches, in which I purpose giving a fuller account of the results than would be consistent with propriety here.

771. *Hydro-cyanic acid* in solution conducts very badly. The definite proportion of hydrogen (equal to that from water) was set free at the *cathode*, whilst at the *anode* a small quantity of oxygen was evolved and apparently a solution of cyanogen formed. The action altogether corresponded with that on a dilute muriatic or hydriodic acid. When the hydro-cyanic acid was made a better conductor by sulphuric acid, the same results occurred.

Cyanides.—With a solution of the cyanide of potassium, the result was precisely the same as with a chloride or iodide. No oxygen was evolved at the positive electrode, but a brown solution formed there. For the reasons given when speaking of the chlorides (766.), and because a fused cyanide of potassium evolves cyanogen at the positive electrode*, I incline to believe that the cyanide in solution is *directly* decomposed.

772. *Ferro-cyanic acid* and the *ferro-cyanides*, as also *sulpho-cyanic acid* and the *sulpho-cyanides*, presented results corresponding with those just described (771.).

773. *Acetic acid.* Glacial acetic acid, when fused (405.), is not decomposed by, nor does it conduct, electricity. On adding a little water to it, still there were no signs of action; on adding more water, it acted slowly and about as water alone would do. Dilute sulphuric acid was added to it in order to make it a better conductor; then the definite proportion of hydrogen was evolved at the *cathode*, and a mixture of oxygen in very deficient quantity, with carbonic acid, and a little carbonic oxide, at the *anode*. Hence it appears that acetic acid is not electrolyzable, but that a portion of it is decomposed by the oxygen evolved at the *anode*, producing secondary results,

* It is a very remarkable thing to see carbon and nitrogen in this case determined powerfully towards the positive surface of the voltaic battery; but it is perfectly in harmony with the theory of electro-chemical decomposition which I have advanced.

varying with the strength of the acid, the intensity of the current, and other circumstances.

774. *Acetates.*—One of these has been referred to already, as affording only secondary results relative to the acetic acid (749.). With many of the metallic acetates the results at both electrodes are secondary (746. 750.).

Acetate of soda fused and anhydrous is directly decomposed, being, as I believe, a true electrolyte, and evolving soda and acetic acid at the *cathode* and *anode*. These, however, have no sensible duration, but are immediately resolved into other substances; charcoal, sodiuretted hydrogen, &c., being set free at the former, and as far as I could judge under the circumstances, acetic acid mingled with carbonic oxide, carbonic acid, &c., at the latter.

775. *Tartaric acid.*—Pure solution of tartaric acid is almost as bad a conductor as pure water. On adding sulphuric acid to it, it conducted well, the results at the positive electrode being primary or secondary in different proportions, according to variations in the strength of the acid and the power of the electric current (752.). Alkaline tartrates gave a large proportion of secondary results at the positive electrode. The hydrogen at the negative electrode remained constant unless certain metallic salts were used.

776. Solutions of salts containing other vegetable acids, as the benzoates; of sugar, gum, &c., dissolved in dilute sulphuric acid; of resin, albumen, &c., dissolved in alkalies, were in turn submitted to the electrolytic power of the voltaic current. In all these cases, secondary results to a greater or smaller extent were produced at the positive electrode.

777. In concluding this division of these Researches, it cannot but occur to the mind that the final result of the action of the electric current upon substances placed between the electrodes, instead of being simple may be very complicated. There are two modes by which these substances may be decomposed, either by the direct force of the electric current, or by the action of bodies which that current may evolve. There are also two modes by which new compounds may be formed, i. e. by combination of the evolving substances whilst in their nascent state (658.), directly with the matter of the electrode; or else their combination with those bodies, which being contained in, or associated with, the decomposing conductor, are necessarily present at the *anode* and *cathode*. The complexity is rendered still greater by the circumstance that two or more of these actions may occur simultaneously, and also in variable proportions to each other. But it may in a great measure be resolved by attention to the principles already laid down (747.).

778. When *aqueous* solutions of bodies are used, secondary results are exceedingly frequent. Even when the water is not present in large quantity, but is merely that of combination, still secondary results often ensue: for instance, it is very possible that in Sir HUMPHRY DAVY's decomposition of the hydrates of potassa and soda, a part of the potassium produced was the result of a secondary action. Hence, also, a frequent

cause for the disappearance of the oxygen and hydrogen which would otherwise be evolved: and when hydrogen does *not* appear at the *cathode* in an *aqueous solution*, it perhaps always indicates that a secondary action has taken place there. No exception to this rule has as yet occurred to my observation.

779. Secondary actions are *not confined to aqueous solutions*, or cases where water is present. For instance, various chlorides acted upon, when fused (402.), by platina electrodes, have the chlorine determined electrically to the *anode*. In many cases, as with the chlorides of lead, potassium, barium, &c., the chlorine acts on the platina and forms a compound with it, which dissolves; but when protochloride of tin is used, the chlorine at the *anode* does not act upon the platina, but upon the chloride already there, forming a perchloride which rises in vapour (790. 804.). These are, therefore, instances of secondary actions of both kinds, produced in bodies containing no water.

780. The production of boron from fused borax (402. 417.) is also a case of secondary action; for boracic acid is not decomposable by electricity (408.), and it was the sodium evolved at the *cathode* which, reacting on the boracic acid around it, took oxygen from it and set boron free in the experiments formerly described.

781. Secondary actions have already, in the hands of M. Becquerel, produced many interesting results in the formation of compounds; some of them new, others imitations of those occurring naturally*. It is probable they may prove equally interesting in an opposite direction, i. e. as affording cases of analytic decomposition. Much information regarding the composition, and perhaps even the arrangement of the particles of such bodies as the vegetable acids and alkalies, and organic compounds generally, will probably be obtained by submitting them to the action of nascent oxygen, hydrogen, chlorine, &c., at the electrodes; and the action seems the more promising, because of the thorough command which we possess over attendant circumstances, such as the strength of the current, the size of the electrodes, the nature of the decomposing conductor, its strength, &c., all of which may be expected to have their corresponding influence upon the final result.

782. It is to me a great satisfaction that the extreme variety of secondary results have presented nothing opposed to the doctrine of a constant and definite electro-chemical action, to the particular consideration of which I shall now proceed.

¶ vii. *On the definite nature and extent of Electro-chemical Decomposition.*

783. In the third series of these Researches, after proving the identity of electricities derived from different sources, and showing, by actual measurement, the extraordinary quantity of electricity evolved by a very feeble voltaic arrangement (371. 376.), I announced a law, derived from experiment, which seemed to me of the utmost importance to the science of electricity in general, and that branch of it denominated electro-chemistry in particular. The law was expressed thus: *The chemical power of*

* Annales de Chimie, tom. xxxv. p. 113.

a current of electricity is in direct proportion to the absolute quantity of electricity which passes (377.).

784. In the further progress of the successive investigations, I have had frequent occasion to refer to the same law, occasionally in circumstances offering powerful corroboration of its truth (456. 504. 505.); and the present series already supplies numerous new cases in which it holds good (704. 722. 726. 732.). It is now my object to consider this great principle more closely, and to develope some of the consequences to which it leads. That the evidence for it may be the more distinct and applicable, I shall quote cases of decomposition subject to as few interferences from secondary results as possible, effected upon bodies very simple, yet very definite in their nature.

785. In the first place, I consider the law as so fully established with respect to the decomposition of *water*, and under so many circumstances which might be supposed, if anything could, to exert an influence over it, that I may be excused entering into further detail respecting that substance, or even summing up the results here (732). I refer, therefore, to the whole of the subdivision of this series of Researches which contains the account of the *volta-electrometer*.

786. In the next place, I also consider the law as established with respect to *muriatic acid* by the experiments and reasoning already advanced, when speaking of that substance, in the subdivision respecting primary and secondary results (758, &c.).

787. I consider the law as established also with regard to *hydriodic acid* by the experiments and considerations already advanced in the preceding division of this series of Researches (767. 768.).

788. Without speaking with the same confidence, yet from the experiments described, and many others not described, relating to hydro-fluoric, hydro-cyanic, ferro-cyanic, and sulpho-cyanic acids (770. 771. 772.), and from the close analogy which holds between these bodies and the hydro-acids of chlorine, iodine, bromine, &c., I consider these also as coming under subjection to the law, and assisting to prove its truth.

789. In the preceding cases, except the first, the water is believed to be inactive; but to avoid any ambiguity arising from its presence, I sought for substances from which it should be absent altogether; and, taking advantage of the law of conduction already developed (380. &c.), soon found abundance, amongst which *protochloride of tin* was first subjected to decomposition in the following manner. A piece of platina wire had one extremity coiled up into a small knob, and having been carefully weighed, was sealed hermetically into a piece of bottle-glass tube, so that the knob should be at the bottom of the tube within (fig. 13.). The tube was suspended by a piece of platina wire, so that the heat of a spirit-lamp could be applied to it. Recently fused protochloride of tin was introduced in sufficient quantity to occupy, when melted, about one half of the tube; the wire of the tube was connected with a volta-electrometer (711.), which was itself connected with the negative end of a voltaic battery; and a platina wire connected with the positive end of the same battery was dipped into the

fused chloride in the tube; being, however, so bent, that it could not by any shake of the hand or apparatus touch the negative electrode at the bottom of the vessel. The whole arrangement is delineated fig. 14.

790. Under these circumstances the chloride of tin was decomposed: the chlorine evolved at the positive electrode formed bichloride of tin (779.), which passed away in fumes, and the tin evolved at the negative electrode combined with the platina, forming an alloy, fusible at the temperature to which the tube was subjected, and therefore never occasioning metallic communication entirely through the decomposing chloride. When the experiment had been continued so long as to yield a reasonable quantity of gas in the volta-electrometer, the battery connexion was broken, the positive electrode removed, and the tube and remaining chloride allowed to cool. When cold, the tube was broken open, the rest of the chloride and the glass being easily separable from the platina wire and its button of alloy. The latter when washed was then reweighed, and the increase gave the weight of the tin reduced.

791. I will give the particular results of one experiment, in illustration of the mode adopted in this and others, the results of which I shall have occasion to quote. The negative electrode weighed at first 20 grains; after the experiment it, with its button of alloy, weighed 23·2 grains. The tin evolved by the electric current at the *cathode* weighed, therefore, 3·2 grains. The quantity of oxygen and hydrogen collected in the volta-electrometer = 3·85 cubic inches. As 100 cubic inches of oxygen and hydrogen, in the proportions to form water, may be considered as weighing 12·92 grains, the 3·85 cubic inches would weigh 0·49742 of a grain; that being, therefore, the weight of water decomposed by the same electric current as was able to decompose such weight of protochloride of tin as could yield 3·2 grains of metal. Now 0·49742 : 3·2 :: 9 the equivalent of water is to 57·9, which should therefore be the equivalent of tin, if the experiment had been made without error, and if the electro-chemical decomposition *is in this case also definite*. In some chemical works 58 is given as the chemical equivalent of tin, in others 57·9. Both are so near to the result of the experiment, and the experiment itself is so subject to slight causes of variation (as from the absorption of gas in the volta-electrometer (716.), &c.), that the numbers leave little doubt of the applicability of the *law of definite action* in this and all similar cases of electro-decomposition.

792. It is not often I have obtained an accordance in numbers so near as that I have just quoted. Four experiments were made on the protochloride of tin, the quantities of gas evolved in the volta-electrometer being from 2·05 to 10·29 cubic inches. The average of the four experiments gave 58·53 as the electro-chemical equivalent for tin.

793. The chloride remaining after the experiment, was pure protochloride of tin; and no one can doubt for a moment that the equivalent of chlorine had been evolved at the *anode*, and having formed bichloride of tin as a secondary result, had passed away.

794. *Chloride of lead* was experimented upon in a manner exactly similar, except that a change was made in the nature of the positive electrode; for as the chlorine evolved at the *anode* forms no perchloride of lead, but acts directly upon the platina, if that metal be used, it produces a solution of chloride of platina in the chloride of lead; in consequence of which a portion of platina can pass to the *cathode*, and will produce a vitiated result. I therefore sought for, and found in plumbago, another substance, which could be used safely as the positive electrode in such bodies as chlorides, iodides, &c. The chlorine or iodine does not act upon it, but is evolved in the free state; and the plumbago has no reaction, under the circumstances, upon the fused chloride or iodide in which it is plunged. Even if a few particles of plumbago should separate by the heat or the mechanical action of the evolved gas, they can do no harm in the chloride.

795. The mean of three experiments gave the number of 100·85 as the equivalent for lead. The chemical equivalent is 103·5. The deficiency in my experiments I attribute to the solution of part of the gas (716.) in the volta-electrometer; but the results leave no doubt on my mind that both the lead and the chlorine are, in this case, evolved in *definite quantities* by the action of a given quantity of electricity (814. &c.).

796. *Chloride of antimony.*—It was in endeavouring to obtain the electro-chemical equivalent of antimony from the chloride that I found reasons for the statement I have made respecting the presence of water in it in an earlier part of these Researches (690. 693. &c.).

797. I endeavoured to experiment upon the *oxide of lead* obtained by fusion and ignition of the nitrate in a platina crucible, but found great difficulty, from the high temperature required for perfect fusion, and the powerful fluxing qualities of the substance. Green glass tubes repeatedly failed. I at last fused the oxide in a small porcelain crucible, heated fully in a charcoal fire; and as it was essential that the evolution of the lead at the *cathode* should take place beneath the surface, the negative electrode was guarded by a green glass tube, fused around it in such a manner as to expose only the knob of platina at the lower end (fig. 15.), so that it could be plunged beneath the surface, and thus exclude contact of air or oxygen with the lead reduced there. A platina wire was employed for the positive electrode, that metal not being subject to any action from the oxygen evolved against it. The arrangement is given fig. 16.

798. In an experiment of this kind the equivalent for the lead came out 93·17, which is very much too small. This, I believe, was because of the small interval between the positive and negative electrodes in the oxide of lead, so that it was not unlikely that some of the froth and bubbles formed by the oxygen at the *anode* should occasionally even touch the lead reduced at the *cathode*, and re-oxidize it. When I endeavoured to correct this by having more litharge, the greater heat required to keep it all fluid caused a quicker action on the crucible, which was soon eaten through, and the experiment stopped.

799. In one experiment of this kind I used borate of lead (408. 673.). It evolves lead, under the influence of the electric current, at the *anode*, and oxygen at the *cathode*; and as the boracic acid is not either directly (408.) or incidentally decomposed during the operation, I expected a result dependent on the oxide of lead. The borate is not so violent a flux as the oxide, but it requires a higher temperature to make it quite liquid; and if not very hot, the bubbles of oxygen cling to the positive electrode, and retard the transfer of electricity. The number for lead came out 101·29, which is so near to 103·5 as to show that the action of the current had been definite.

800. *Oxide of bismuth.*—I found this substance required too high a temperature, and acted too powerfully as a flux, to allow of any experiment being made on it, without the application of more time and care than I could give at present.

801. The ordinary *protoxide of antimony*, which consists of one proportional of metal and one and a half of oxygen, was subjected to the action of the electric current in a green glass tube (789.), surrounded by a jacket of platina foil, and heated in a charcoal fire. The decomposition began and proceeded very well at first, apparently indicating, according to the general law (679. 697.), that this substance was one containing such elements and in such proportions as made it amenable to the power of the electric current. This effect I have already given reasons for supposing may be due to the presence of a true protoxide, consisting of single proportionals (696. 693.). The action soon diminished, and finally ceased, because of the formation of a higher oxide of the metal at the positive electrode. This compound, which was probably the peroxide, being infusible and insoluble in the protoxide, formed a crystalline crust around the positive electrode; and thus insulating it, prevented the transmission of the electricity. Whether if it had been fusible and still immiscible it would have decomposed, is doubtful, because of its departure from the required composition (697.). It was a very natural secondary product at the positive electrode (779.). On opening the tube it was found that a little antimony had been separated at the negative electrode; but the quantity was too small to allow of any quantitative result being obtained.

802. *Iodide of lead.*—This substance can be experimented with in tubes heated by a spirit-lamp (789.); but I obtained no good results from it, whether I used positive electrodes of platina or plumbago. In two experiments the numbers for the lead came out only 75·46 and 73·45, instead of 103·5. This I attribute to the formation of a periodide at the positive electrode, which dissolving in the mass of liquid iodide, came in contact with the lead evolved at the negative electrode, and dissolved part of it, becoming itself again protiodide. Such a periodide does exist; and it is very rarely that the iodide of lead formed by precipitation, and well washed, can be fused without evolving much iodine, from the presence of this percompound; nor does crystallization from its hot aqueous solution free it from this substance. Even when a little of the protiodide and iodine are merely rubbed together in a mortar, a portion of the periodide is formed. And though it is decomposed by being fused and heated

to dull redness for a few minutes, and the whole reduced to protiodide, yet that is not at all opposed to the possibility, that a little of that which is formed in great excess of iodine at the *anode*, should be carried by the rapid currents in the liquid into contact with the *cathode*.

803. This view of the results was strengthened by a third experiment, where the space between the electrodes was increased to one third of an inch; for now the interfering effects were much diminished, and the number of the lead came out 89·04; and it was fully confirmed by the results obtained in the cases of transfer to be immediately described (818.).

The experiments on iodide of lead, therefore, offer no exception to the *general law* under consideration, but, on the contrary, may, from general considerations, be admitted as included in it.

804. *Protiodide of tin.*—This substance, when fused (402.), conducts and is decomposed by the electric current, tin is evolved at the *anode*, and periodide of tin as a secondary result (779. 790.) at the *cathode*. The temperature required for its fusion is too high to allow of the production of any results fit for weighing.

805. *Iodide of potassium* was subjected to electrolytic action in a tube, fig. 13. (789.). The negative electrode was a globule of lead, and I hoped in this way to retain the potassium, and obtain results that could be weighed and compared with the volta-electrometer indication; but the difficulties dependent upon the high temperature required, the action upon the glass, the fusibility of the platina induced by the presence of the lead, and other circumstances, prevented me from obtaining such results. The iodide was decomposed with the evolution of iodine at the *anode*, and of potassium at the *cathode*, as in former cases.

806. In some of these experiments several substances were placed in succession, and decomposed simultaneously by the same electric current: thus, protochloride of tin, chloride of lead, and water, were thus acted on at once. It is needless to say that the results were comparable, the tin, lead, chlorine, oxygen, and hydrogen evolved being definite in quantity and electro-chemical equivalents to each other.

807. Let us turn to another kind of proof of the *definite chemical action of electricity*. If any circumstances could be supposed to exert an influence over the quantity of the matters evolved during electrolytic action, one would expect them to be present when electrodes of different substances, and possessing very different chemical affinities for the evolving bodies, were used. Platina has no power in dilute sulphuric acid of combining with the oxygen at the *anode*, though the latter be evolved in the nascent state against it. Copper, on the other hand, immediately unites to the oxygen, as the electric current sets it free from the hydrogen; and zinc is not only able to combine with it, but can, without any help from the electricity, abstract it directly from the water, at the same time setting torrents of hydrogen free. Yet in cases where these three substances were used as the positive electrodes in three similar portions of the same dilute sulphuric acid, specific gravity 1·336, precisely the same quantity of water

was decomposed by the electric current, and precisely the same quantity of hydrogen set free at the *cathodes* of the three solutions.

808. The experiment was made thus. Portions of the dilute sulphuric acid were put into three basins. Three volta-electrometer tubes, of the form figg. 5, 7. were filled with the same acid, and one inverted in each basin (707.). A zinc plate, connected with the positive end of a voltaic battery, was dipped into the first basin, forming the positive electrode there, the hydrogen, which was abundantly evolved from it by the direct action of the acid, being allowed to escape. A copper plate, which dipped into the acid of the second basin, was connected with the negative electrode of the *first* basin; and a platina plate, which dipped into the acid of the third basin, was connected with the negative electrode of the *second* basin. The negative electrode of the third basin was connected with a volta-electrometer (711.), and that with the negative end of the voltaic battery.

809. Immediately that the circuit was complete, the *electro-chemical action* commenced in all the vessels. The hydrogen still rose in, apparently, undiminished quantities from the positive zinc electrode in the first basin. No oxygen was evolved at the positive copper electrode in the second basin, but a sulphate of copper was formed there; whilst in the third basin the positive platina electrode evolved pure oxygen gas, and was itself unaffected. But in *all* the basins the hydrogen liberated at the *negative* platina electrodes was the *same in quantity*, and the same with the volume of hydrogen evolved in the volta-electrometer, showing that in all the vessels the current had decomposed an equal quantity of water. In this trying case, therefore, the *chemical action of electricity* proved to be *perfectly definite.*

810. A similar experiment was made with muriatic acid diluted with its bulk of water. The three positive electrodes were zinc, silver, and platina; the first being able to separate and combine with the chlorine *without* the aid of the current; the second combining with the chlorine only after the current had set it free; and the third rejecting almost the whole of it. The three negative electrodes were, as before, platina plates fixed within glass tubes. In this experiment, as in the former, the quantity of hydrogen evolved at the *cathodes* was the same for all, and the same as the hydrogen evolved in the volta-electrometer. I have already given my reasons for believing that in these experiments it is the muriatic acid which is directly decomposed by the electricity (764.); and the results prove that the quantities so decomposed are *perfectly definite* and proportionate to the quantity of electricity which has passed.

811. In this experiment the chloride of silver formed in the second basin retarded the passage of the current of electricity, by virtue of the law of conduction before described (394.), so that it had to be cleaned off four or five times during the course of the experiment; but this caused no difference between the results of that vessel and the others.

812. Charcoal was used as the positive electrode in both sulphuric and muriatic acids (808. 810.); but this change produced no variation of the results. A zinc positive

electrode, in sulphate of soda or solution of common salt, gave the same constancy of operation.

813. Experiments of a similar kind were then made with bodies altogether in a different state, i. e. with *fused* chlorides, iodides, &c. I have already described an experiment with fused chloride of silver, in which the electrodes were of metallic silver, the one rendered negative becoming increased and lengthened by the addition of metal, whilst the other was dissolved and eaten away by its abstraction. This experiment was repeated, two weighed pieces of silver wire being used as the electrodes, and a volta-electrometer included in the circuit. Great care was taken to withdraw the negative electrode so regularly and steadily that the crystals of reduced silver should not form a *metallic* communication beneath the surface of the fused chloride. On concluding the experiment the positive electrode was re-weighed, and its loss ascertained. The mixture of chloride of silver, and metal, withdrawn in successive portions at the negative electrode, was digested in solution of ammonia, to remove the chloride, and the metallic silver remaining also weighed: it was the reduction at the *cathode*, and exactly equalled the solution at the *anode*; and each portion was as nearly as possible the equivalent to the water decomposed in the volta-electrometer.

814. The infusible condition of the silver at the temperature used, and the length and ramifying character of its crystals, render the above experiment difficult to perform, and uncertain in its results. I therefore wrought with a chloride of lead, using a green glass tube, formed as in fig. 17. A weighed platina wire was fused into the bottom of a small tube, as before described (789.). The tube was then bent to an angle, at about half an inch distance from the closed end; and the part between the angle and the extremity being softened, was forced upward, as in the figure, so as to form a bridge, or rather separation, producing two little depressions or basins *a*, *b*, within the tube. This arrangement was suspended by a platina wire, as before, so that the heat of a spirit-lamp could be applied to it, such inclination being given to it as would allow all air to escape during the fusion of the chloride of lead. A positive electrode was then provided, by binding up the end of a platina wire into a knob, and fusing about twenty grains of metallic lead on to it, in a small closed tube of glass, which was afterwards broken away. Being so furnished, the wire with its knob was weighed, and the weight recorded.

815. Chloride of lead was now introduced into the tube, and carefully fused. The leaded electrode was also introduced; after which the metal, at its extremity, soon melted. In this state of things the tube was filled up to *c* with melted chloride of lead; the end of the electrode to be rendered negative was in the basin *b*, and the electrode of melted lead was retained in the basin *a*, and, by connexion with the proper conducting wire of a voltaic battery, was rendered positive. A volta-electrometer was included in the circuit.

816. Immediately upon the completion of the communication with the voltaic battery, the current passed, and decomposition proceeded. No chlorine was evolved at

the positive electrode; but as the fused chloride was transparent, a button of alloy could be observed gradually forming and increasing in size at *b*, whilst the lead at *a* could also be seen gradually to diminish. After a time, the experiment was stopped; the tube allowed to cool, and broken open; the wires, with their buttons, cleaned and weighed; and their change in weight compared with the indication of the *volta-electrometer*.

817. In this experiment the positive electrode had lost just as much lead as the negative one had gained (795.), and the loss or gain was very nearly the equivalent of the water decomposed in the volta-electrometer, giving for lead the number 101·5. It is therefore evident, in this instance, that causing a *strong affinity*, or *no affinity*, for the substance evolved at the *anode*, to be active during the experiment (807.), produces no variation in the definite action of the electric current.

818. A similar experiment was then made with iodide of lead, and in this manner all confusion from the formation of a periodide avoided (803.). No iodine was evolved during the whole action, and finally the loss of lead at the *anode* was the same as the gain at the *cathode*, the equivalent number, by comparison with the result in the volta-electrometer, being 103·5.

819. Then protochloride of tin was subjected to the electric current in the same manner, using, of course, a tin positive electrode. No bichloride of tin was now formed (779. 790.). On examining the two electrodes, the positive had lost precisely as much as the negative had gained; and by comparison with the volta-electrometer, the number for tin came out 59.

820. It is quite necessary in these and similar experiments to examine the interior of the bulbs of alloy at the ends of the conducting wires; for occasionally, and especially with those which have been positive, they are cavernous, and contain portions of the chloride or iodide used, which must be removed before the final weight is ascertained. This is more usually the case with lead than tin.

821. All these facts combine into, I think, an irresistible mass of evidence, proving the truth of the important proposition which I at first laid down, namely, *that the chemical power of a current of electricity is in direct proportion to the absolute quantity of electricity which passes* (377. 783.). They prove, too, that this is not merely true with one substance, as water, but generally with all electrolytic bodies; and, further, that the results obtained with any *one substance* do not merely agree amongst themselves, but also with those obtained from *other substances*, the whole combining together into *one series of definite electro-chemical actions* (505.). I do not mean to say that no exceptions will appear: perhaps some may arise, especially amongst substances existing only by weak affinity; but I do not expect that any will seriously disturb the result announced. If, in the well considered, well examined, and, I may surely say, well ascertained doctrines of the definite nature of ordinary chemical affinity, such exceptions occur, as they do in abundance, yet, without being allowed to disturb our minds as to the general conclusion, they ought also to be allowed if they should

present themselves at this, the opening of a new view of electro-chemical action; not being held up as obstructions to those who may be engaged in rendering that view more and more perfect, but laid aside for a while, in hopes that their perfect and consistent explanation will finally appear.

---

822. The doctrine of *definite electro-chemical action* just laid down, and, I believe, established, leads to some new views of the relations and classifications of bodies associated with or subject to this action. Some of these I shall proceed to consider.

823. In the first place, compound bodies may be separated into two great classes, namely, those which are decomposable by the electric current, and those which are not. Of the latter, some are conductors, others non-conductors, of voltaic electricity*. The former do not depend for their decomposability, upon the nature of their elements only; for, of the same two elements, bodies may be formed, of which one shall belong to one class and another to the other class; but probably on the proportions also (697.). It is further remarkable, that with very few, if any, exceptions (414. 691.), these decomposable bodies are exactly those governed by the remarkable law of conduction I have before described (394.); for that law does not extend to the many compound fusible substances that are excluded from this class. I propose to call bodies of this, the decomposable class, *Electrolytes* (664.).

824. Then, again, the substances into which these divide, under the influence of the electric current, form an exceedingly important general class. They are combining bodies; are directly associated with the fundamental parts of the doctrine of chemical affinity; and have each a definite proportion, in which they are always evolved during electrolytic action. I have proposed to call these bodies generally *ions*, or particularly *anions* and *cations*, according as they appear at the *anode* or *cathode* (665.); and the numbers representing the proportions in which they are evolved *electro-chemical equivalents*. Thus hydrogen, oxygen, chlorine, iodine, lead, tin, are *ions*; the three former are *anions*, the two metals are *cations*, and 1, 8, 36, 125, 104, 58, are their *electro-chemical equivalents* nearly.

825. A summary of certain points already ascertained respecting *electrolytes, ions*, and *electro-chemical equivalents*, may be given in the following general form of propositions, without, I hope, including any serious error.

826. i. A single *ion*, i. e. one not in combination with another, will have no tendency to pass to either of the electrodes, and will be perfectly indifferent to the passing current, unless it be itself a compound of more elementary *ions*, and so subject to actual decomposition. Upon this fact is founded much of the proof adduced in favour of the new theory of electro-chemical decomposition, which I put forth in a former series of these Researches (518. &c.).

827. ii. If one *ion* be combined in right proportions (697.) with another strongly

* I mean here by voltaic electricity, merely electricity from a most abundant source, but having very small intensity.

opposed to it in its ordinary chemical relations, i. e. if an *anion* be combined with a *cation*, then both will travel, the one to the *anode*, the other to the *cathode*, of the decomposing body (530. 542. 547.).

828. iii. If, therefore, an *ion* pass towards one of the electrodes, another *ion* must also be passing simultaneously to the other electrode, although, from secondary action, it may not make its appearance (743.).

829. iv. A body decomposable directly by the electric current, i. e. an *electrolyte*, must consist of two *ions*, and must also render them up during the act of decomposition.

830. v. There is but one *electrolyte* composed of the same two elementary *ions*; at least such appears to be the fact (697.), dependent upon a law, that *only single electro-chemical equivalents of elementary ions can go to the electrodes, and not multiples.*

831. vi. A body not decomposable when alone, as boracic acid, is not directly decomposable by the electric current when in combination (780.). It may act as an *ion*, going wholly to the *anode* or *cathode*, but does not yield up its elements, except occasionally by a secondary action. Perhaps it is superfluous for me to point out that this proposition has *no relation* to such cases as that of water, which, by the presence of other bodies, is rendered a better conductor of electricity, and *therefore* is more freely decomposed.

832. vii. The nature of the substance of which the electrode is formed, provided it be a conductor, causes no difference in the electro-decomposition, either in kind or degree (807. 813.); but it seriously influences, by secondary action (744.), the state in which the *ions* finally appear. Advantage may be taken of this principle in combining and collecting such *ions* as, if evolved in their free state, would be unmanageable *.

833. viii. A substance which, being used as the electrode, can combine altogether with the *ion* evolved against it, is also, I believe, an *ion*, and combines, in such cases, in the quantity represented by its *electro-chemical equivalent.* All the experiments I have made agree with this view; and it seems to me, at present, to result as a necessary consequence. Whether, in the secondary actions that take place, where the *ion* acts, not upon the matter of the electrode, but on that which is around it in the liquid (744.), the same consequence follows, will require more extended investigation to determine.

834. ix. Compound *ions* are not necessarily composed of electro-chemical equivalents of simple *ions*. For instance, sulphuric acid, boracic acid, phosphoric acid, are *ions*, but not *electrolytes*, i. e. not composed of electro-chemical equivalents of simple *ions*.

* It will often happen that the electrodes used may be of such a nature as, with the fluid in which they are immersed, to produce an electric current, either according with or opposing that of the voltaic arrangement used, and in this way, or by direct chemical action, may sadly disturb the results. Still, in the midst of all these confusing effects, the electric current, which actually passes in any direction through the decomposing body, will produce its own definite electrolytic action.

835. x. Electro-chemical equivalents are always consistent; i. e. the same number which represents the equivalent of a substance A when it is separating from a substance B, will also represent A when separating from a third substance C. Thus, 8 is the electro-chemical equivalent of oxygen, whether separating from hydrogen, or tin, or lead; and 103·5 is the electro-chemical equivalent of lead, whether separating from oxygen, or chlorine, or iodine.

836. xi. Electro-chemical equivalents coincide, and are the same, with ordinary chemical equivalents.

837. By means of experiment and the preceding propositions, a knowledge of *ions* and their electro-chemical equivalents may be obtained in various ways.

838. In the first place, they may be determined directly, as has been done with hydrogen, oxygen, lead, and tin, in the numerous experiments already quoted.

839. In the next place, from propositions ii. and iii., may be deduced the knowledge of many other *ions*, and also their equivalents. When chloride of lead was decomposed, platina being used for both electrodes (395.), there could remain no more doubt that chlorine was passing to the *anode*, although it combined with the platina there, than when the positive electrode, being of plumbago (794.), allowed its evolution in the free state; neither could there, in either case, remain any doubt, that for every 103·5 parts of lead evolved at the *cathode*, 36 parts of chlorine were evolved at the *anode*, for the remaining chloride of lead was unchanged. So also when in a metallic solution one volume of oxygen, or a secondary compound containing that proportion, appeared at the *anode*, no doubt could arise that hydrogen, equivalent to two volumes, had been determined to the *cathode*, although, by a secondary action, it had been employed in reducing oxides of lead, copper, or other metals, to the metallic state. In this manner, then, we learn from the experiments already described in these Researches, that chlorine, iodine, bromine, fluorine, calcium, potassium, strontium, magnesium, manganese, &c., are *ions*, and that their *electro-chemical equivalents* are the same as their *ordinary chemical equivalents*.

840. Propositions iv. and v. extend our means of gaining information. For if a body of known chemical composition is found to be decomposable, and the nature of the substance evolved as a primary or even a secondary result (743. 777.) at one of the electrodes, be ascertained, the electro-chemical equivalent of that body may be deduced from the known constant composition of the substance evolved. Thus, when fused protiodide of tin is decomposed by the voltaic current (804.), the conclusion may be drawn, that both the iodine and tin are *ions*, and that the proportions in which they combine in the fused compound express their electro-chemical equivalents. Again, with respect to the fused iodide of potassium (805.), it is an electrolyte; and the chemical equivalents will also be the electro-chemical equivalents.

841. If proposition viii. sustain extensive experimental investigation, then it will not only help to confirm the results obtained by the use of the other propositions, but will give abundant original information of its own.

842. In many instances, the *secondary results* obtained by the action of the evolving *ion* on the substances present in the surrounding liquid or solution, will give the electro-chemical equivalent. Thus, in the solution of acetate of lead, and, as far as I have gone, in other proto-salts subjected to the reducing action of the nascent hydrogen at the *cathode*, the metal precipitated has been in the same quantity as if it had been a primary product, (provided no free hydrogen escaped there,) and therefore gave as accurately the number representing its electro-chemical equivalent.

843. Upon this principle it is that secondary results may occasionally be used as measurers of the volta-electric current (706. 740.); but there are not many metallic solutions that answer this purpose well: for unless the metal is easily precipitated, hydrogen will be evolved at the *cathode* and vitiate the result. If a soluble peroxide is formed at the *anode*, or if the precipitated metal crystallize across the solution and touch the positive electrode, similar vitiated results are obtained. I expect to find in some vegetable salts, as the acetates of mercury and zinc, solutions favourable for this use.

844. After the first experimental investigations to establish the definite chemical action of electricity, I have not hesitated to apply the more strict results of chemical analysis to correct the numbers obtained as electrolytical results. This, it is evident, may be done in a great number of cases, without using too much liberty towards the due severity of scientific research. The series of numbers representing electro-chemical equivalents must, like those expressing the ordinary equivalents of chemically acting bodies, remain subject to the continual correction of experiment and sound reasoning.

845. I give the following brief Table of *ions* and their electro-chemical equivalents, rather as a specimen of a first attempt than as anything that can supply the want which must very quickly be felt, of a full and complete tabular account of this class of bodies. Looking forward to such a table as of extreme utility (if well constructed) in developing the intimate relation of ordinary chemical affinity to electrical actions, and identifying the two, not to the imagination merely, but to the conviction of the senses and a sound judgement, I may be allowed to express a hope, that the endeavour will always be to make it a table of *real*, and not *hypothetical*, electro-chemical equivalents; for we shall else overrun the facts, and lose all sight and consciousness of the knowledge lying directly in our path.

846. The equivalent numbers do not profess to be exact, and are taken almost entirely from the chemical results of other philosophers in whom I could repose more confidence, as to these points, than in myself.

847. TABLE OF IONS.

Anions.

Oxygen	8	Cyanogen	26	Phosphoric acid	35·7	Citric acid	58
Chlorine	35·5	Sulphuric acid	40	Carbonic acid	22	Oxalic acid	36
Iodine	126	Selenic acid	64	Boracic acid	24	Sulphur (?)	16
Bromine	78·3	Nitric acid	54	Acetic acid	51	Selenium (?)	
Fluorine	18·7	Chloric acid	75·5	Tartaric acid	66	Sulpho-cyanogen	

Cations.

Hydrogen	1	Tin	57·9	Mercury	200	Strontia	51·8
Potassium	39·2	Lead	103·5	Silver	108	Lime	28·5
Sodium	23·3	Iron	28	Platina	98·6?	Magnesia	20·7
Lithium	10	Copper	31·6	Gold	(?)	Alumina	(?)
Barium	68·7	Cadmium	55·8	———		Protoxides generally.	
Strontium	43·8	Cerium	46	Ammonia	17	Quinia	171·6
Calcium	20·5	Cobalt	29·5	Potassa	47·2	Cinchona	160
Magnesium	12·7	Nickel	29·5	Soda	31·3	Morphia	290
Manganese	27·7	Antimony	64·6?	Lithia	18	Vegeto-alkalies generally.	
Zinc	32·5	Bismuth	71	Baryta	76·7		

848. This Table might be further arranged into groups of such substances as either act with, or replace, each other. Thus, for instance, acids and bases act in relation to each other; but they do not act in association with oxygen, hydrogen, or elementary substances. There is indeed little or no doubt that, when the electrical relations of the particles of matter come to be closely examined, this division must be made. The simple substances, with cyanogen, sulpho-cyanogen, and one or two other compound bodies, will probably form the first group; and the acids and bases, with such analogous compounds as may prove to be *ions*, the second group. Whether these will include all *ions*, or whether a third class of more complicated results will be required, must be decided by future experiments.

849. It is *probable* that all our present elementary bodies are *ions*, but that is not as yet certain. There are some, such as carbon, phosphorus, nitrogen, silicon, boron, alumium, the right of which to the title of *ion* it is desirable to decide as soon as possible. There are also many compound bodies, and amongst them alumina and silica, which it is desirable to class immediately by unexceptionable experiments. It is also *possible*, that all combinable bodies, compound as well as simple, may enter into the class of *ions*; but at present it does not seem to me probable. Still the experimental evidence I have is so small in proportion to what must gradually accumulate around, and bear upon, this point, that I am afraid to give a strong opinion upon it.

850. I think I cannot deceive myself in considering the doctrine of definite electro-chemical action as of the utmost importance. It touches by its facts more directly and closely than any former fact, or set of facts, have done, upon the beautiful idea, that ordinary chemical affinity is a mere consequence of the electrical attractions of the particles of different kinds of matter; and it will probably lead us to the means by which we may enlighten that which is at present so obscure, and either fully demonstrate the truth of the idea, or develope that which ought to replace it.

851. A very valuable use of electro-chemical equivalents will be to decide, in cases of doubt, what is the true chemical equivalent, or definite proportional, or atomic number of a body; for I have such conviction that the power which governs electro-decomposition and ordinary chemical attractions is the same; and such confidence in the overruling influence of those natural laws which render the former definite, as to

feel no hesitation in believing that the latter must submit to them also. Such being the case, I can have no doubt that, assuming hydrogen as 1, and dismissing small fractions for the simplicity of expression, the equivalent number or atomic weight of oxygen is 8, of chlorine 36, of bromine 78·4, of lead 103·5, of tin 59, &c., notwithstanding that a very high authority doubles several of these numbers.

§ 13. *On the absolute quantity of Electricity associated with the particles or atoms of Matter.*

852. The theory of definite electrolytical or electro-chemical action appears to me to touch immediately upon the *absolute quantity* of electricity or electric power belonging to different bodies. It is impossible, perhaps, to speak on this point without committing oneself beyond what present facts will sustain; and yet it is equally impossible, and perhaps would be impolitic, not to reason upon the subject. Although we know nothing of what an atom is, yet we cannot resist forming some idea of a small particle, which represents it to the mind; and though we are in equal, if not greater, ignorance of electricity, so as to be unable to say whether it is a particular matter or matters, or mere motion of ordinary matter, or some third kind of power or agent, yet there is an immensity of facts which justify us in believing that the atoms of matter are in some way endowed or associated with electrical powers, to which they owe their most striking qualities, and amongst them their mutual chemical affinity. As soon as we perceive, through the teaching of DALTON, that chemical powers are, however varied the circumstances in which they are exerted, definite for each body, we learn to estimate the relative degree of force which resides in such bodies: and when upon that knowledge comes the fact, that the electricity, which we appear to be capable of loosening from its habitation for a while, and conveying from place to place, *whilst it retains its chemical force*, can be measured out, and, being so measured, is found to be *as definite in its action* as any of *those portions* which, remaining associated with the particles of matter, give them their *chemical relation*; we seem to have found the link which connects the proportion of that we have evolved to the proportion of that belonging to the particles in their natural state.

853. Now it is wonderful to observe how small a quantity of a compound body is decomposed by a certain portion of electricity. Let us, for instance, consider this and a few other points in relation to water. *One grain* of water acidulated to facilitate conduction, will require an electric current to be continued for three minutes and three quarters of time to effect its decomposition, which current must be powerful enough to retain a platina wire $\frac{1}{104}$ of an inch in thickness*, red hot, in the air during the whole time; and if interrupted anywhere by charcoal points, will produce a very

* I have not stated the length of wire used, because I find by experiment, as would be expected in theory, that it is indifferent. The same quantity of electricity which, passed in a given time, can heat an inch of platina wire of a certain diameter red hot, can also heat a hundred, a thousand, or any length of the same wire to the same degree, provided the cooling circumstances are the same for every part in both cases. This I have proved by the volta-electrometer. I found that whether half an inch or eight inches were retained at one constant

brilliant and constant star of light. If attention be paid to the instantaneous discharge of electricity of tension, as illustrated in the beautiful experiments of Mr. Wheatstone *, and to what I have said elsewhere on the relation of common and voltaic electricity (371. 375.), it will not be too much to say, that this necessary quantity of electricity is equal to a very powerful flash of lightning. Yet we have it under perfect command; can evolve, direct, and employ it at pleasure; and when it has performed its full work of electrolyzation, it has only separated the elements of a single grain of water.

854. On the other hand, the relation between the conduction of the electricity and the decomposition of the water is so close, that one cannot take place without the other. If the water is altered only in that small degree which consists in its having the solid instead of the fluid state, the conduction is stopped, and the decomposition is stopped with it. Whether the conduction be considered as depending upon the decomposition, or not (413. 703.), still the relation of the two functions is equally intimate and inseparable.

855. Considering this close and twofold relation, namely, that without decomposition transmission of electricity does not occur; and, that for a given definite quantity of electricity passed, an equally definite and constant quantity of water or other matter is decomposed; considering also that the agent, which is electricity, is simply employed in overcoming electrical powers in the body subjected to its action; it seems a probable, and almost a natural consequence, that the quantity which passes is the *equivalent* of, and therefore equal to, that of the particles separated; i. e. that if the electrical power which holds the elements of a grain of water in combination, or which makes a grain of oxygen and hydrogen in the right proportions unite into water when they are made to combine, could be thrown into the condition of *a current*, it would exactly equal the current required for the separation of that grain of water into its elements again.

856. This view of the subject gives an almost overwhelming idea of the extraordinary quantity or degree of electric power which naturally belongs to the particles of matter; but it is not inconsistent in the slightest degree with the facts which can be brought to bear on this point. To illustrate this I must say a few words on the voltaic pile †.

temperature of dull redness, equal quantities of water were decomposed in equal times in both cases. When the half-inch was used, only the centre portion of wire was ignited. A fine wire may even be used as a rough but ready regulator of a voltaic current; for if it be made part of the circuit, and the larger wires communicating with it be shifted nearer to or further apart, so as to keep the portion of wire in the circuit sensibly at the same temperature, the current passing through it will be nearly uniform.

* Literary Gazette, 1833, March 1 and 8. Philosophical Magazine, 1833, p. 204. L'Institute, 1833, p. 261.

† By the term voltaic pile, I mean such apparatus or arrangement of metals as up to this time have been called so, and which contain water, brine, acids, or other aqueous solutions or decomposable substances (476.), between their plates. Other kinds of electric apparatus may be hereafter invented, and I hope to construct some not belonging to the class of instruments discovered by Volta.

857. Intending hereafter to apply the results given in this and the preceding series of Researches to a close investigation of the source of electricity in the voltaic instrument, I have refrained from forming any decided opinion on the subject; and without at all meaning to dismiss metallic contact, or the contact of dissimilar substances, being conductors, but not metallic, as if they had nothing to do with the origin of the current, I still am fully of opinion with DAVY, that it is at least continued by chemical action, and that the supply constituting the current is almost entirely from that source.

858. Those bodies which, being interposed between the metals of the voltaic pile, render it active, *are all of them electrolytes* (476.); and it cannot but press upon the attention of every one engaged in considering this subject, that in those bodies (so essential to the pile) decomposition and the transmission of a current are so intimately connected, that one cannot happen without the other. This I have shown abundantly in water, and numerous other cases (402. 476.). If, then, a voltaic trough have its extremities connected by a decomposing body, as water, we shall have a continuous current through the apparatus; and whilst it remains in this state may look at the part where the acid is acting upon the plates, and that where the current is acting upon the water, as the reciprocals of each other. In both parts we have the two conditions *inseparable in such bodies as these*, namely, the passing of a current, and decomposition; and this is as true of the cells in the battery as of the water cell; for no voltaic battery has as yet been constructed in which the chemical action is only that of combination: decomposition is always included, and is, I believe, an essential chemical part.

859. But the difference in the two parts of the connected battery, that is, the decomposing or experimental cell, and the acting cells, is simply this. In the former we urge the current through, but it, apparently of necessity, is accompanied by decomposition: in the latter we cause decompositions by ordinary chemical actions, (which are, however, themselves electrical,) and, as a consequence, have the electrical current; and as the decomposition dependent upon the current is definite in the former case, so is the current associated with the decomposition also definite in the latter (862. &c.).

860. Let us apply this in support of what I have surmised respecting the enormous electric power of each particle or atom of matter (856.). I showed in a former series of these Researches on the relation by measure of common and voltaic electricity, that two wires, one of platina and one of zinc, each one eighteenth of an inch in diameter, placed five sixteenths of an inch apart, and immersed to the depth of five eighths of an inch in acid, consisting of one drop of oil of vitriol and four ounces of distilled water at a temperature of about 60° FAHR., and connected at the other extremities by a copper wire eighteen feet long, and one eighteenth of an inch in thickness, yielded as much electricity in little more than three seconds of time as a Leyden battery charged by thirty turns of a very large and powerful plate electric machine

in full action (371.). This quantity, though sufficient if passed at once through the head of a rat or a cat to have killed it, as by a flash of lightning, was evolved by the mutual action of so small a portion of the zinc wire and water in contact with it, that the loss of weight sustained by either would be inappreciable by our most delicate instruments; and as to the water which could be decomposed by that current, it must have been insensible in quantity, for no trace of hydrogen appeared upon the surface of the platina during those three seconds.

861. What an enormous quantity of electricity, therefore, is required for the decomposition of a single grain of water! We have already seen that it must be in quantity sufficient to sustain a platina wire $\frac{1}{104}$ of an inch in thickness, red hot, in contact with the air for three minutes and three quarters (853.), a quantity which is almost infinitely greater than that which could be evolved by the little standard voltaic arrangement to which I have just referred (860. 371.). I have endeavoured to make a comparison by the loss of weight of such a wire in a given time in such an acid, according to a principle and experiment to be almost immediately described (862.); but the proportion is so high, that I am almost afraid to mention it. It would appear that 800,000 such charges of the Leyden battery as I have referred to above, would be necessary to supply electricity sufficient to decompose a single grain of water; or, if I am right, to equal the quantity of electricity which is naturally associated with the elements of that grain of water, endowing them with their mutual chemical affinity.

862. In further proof of this high electric condition of the particles of matter, and the *identity as to quantity, of that belonging to them with that necessary for their separation,* I will describe an experiment of great simplicity but extreme beauty, when viewed in relation to the evolution of an electric current and its decomposing powers.

863. A dilute sulphuric acid, made by adding about one part by measure of oil of vitriol to thirty parts of water, will act energetically upon a piece of plate zinc in its ordinary and simple state; but, as Mr. Sturgeon has shown*, not at all, or scarcely so, if the surface of the metal has in the first instance been amalgamated; yet the amalgamated zinc will act powerfully with platina as an electromotor, hydrogen being evolved on the surface of the latter metal, as the zinc is oxidized and dissolved. The amalgamation is best effected by sprinkling a few drops of mercury upon the surface of the zinc, the latter being moistened with the dilute acid, and rubbing with the fingers so as to extend the liquid metal over the whole of the surface. Any mercury in excess forming liquid drops upon the zinc, should be wiped off†.

864. Two plates of zinc thus amalgamated were dried and accurately weighed; one, which we will call A, weighed 163·1 grains; the other, to be called B, weighed 148·3

* Recent Experimental Researches, &c., 1830, p. 74, &c.

† The experiment may be made with pure zinc, which, as chemists well know, is but slightly acted upon by dilute sulphuric acid in comparison with ordinary zinc, which during the action is subject to an infinity of voltaic actions. See De la Rive on this subject, Bibliothèque Universelle, 1830, p. 391.

grains. They were about five inches long, and 0·4 of an inch wide. An earthenware pneumatic trough was filled with dilute sulphuric acid, of the strength just described (863.), and a gas jar, also filled with the acid, inverted in it*. A plate of platina of nearly the same length, but about three times as wide as the zinc plates, was put up into this jar. The zinc plate A was also introduced into the jar, and brought in contact with the platina, and at the same moment the plate B was put into the acid of the trough, but out of contact with other metallic matter.

865. Strong action immediately occurred in the jar upon the contact of the zinc and platina plates. Hydrogen gas rose from the platina, and was collected in the jar, but no hydrogen or other gas rose from *either* zinc plate. In about ten or twelve minutes, sufficient hydrogen having been collected, the experiment was stopped; during its progress a few small bubbles had appeared upon plate B, but none upon plate A. The plates were washed in distilled water, dried, and reweighed. Plate B weighed 148·3 grains, as before, having lost nothing by the direct chemical action of the acid. Plate A weighed 154·65 grains, 8·45 grains of it having been oxidized and dissolved during the experiment.

866. The hydrogen gas was next transferred to a water-trough and measured; it amounted to 12·5 cubic inches, the temperature being 52°, and the barometer 29·2 inches. This quantity, corrected for temperature, pressure, and moisture, becomes 12·15453 cubic inches of dry hydrogen at mean temperature and pressure; which, increased by one half for the oxygen that must have gone to the *anode*, i. e. to the zinc, gives 18·232 cubic inches as the quantity of oxygen and hydrogen evolved from the water decomposed by the electric current. According to the estimate of the weight of the mixed gas before adopted (791.), this volume is equal to 2·3535544 grains, which therefore is the weight of water decomposed; and this quantity is to 8·45, the quantity of zinc oxidized, as 9 is to 32·31. Now taking 9 as the equivalent number of water, the number 32·5 is given as the equivalent number of zinc; a coincidence sufficiently near to show, what indeed could not but happen, that for an equivalent of zinc oxidized an equivalent of water must be decomposed†.

867. But let us observe *how* the water is decomposed. It is electrolyzed, i. e. is decomposed voltaically, and not in the ordinary manner (as to appearance) of chemical decompositions; for the oxygen appears at the *anode* and the hydrogen at the *cathode* of the decomposing body, and these were in many parts of the experiment above an inch asunder. Again, the ordinary chemical affinity was not enough under the circumstances to effect the decomposition of the water, as was abundantly proved by the inaction on plate B; the voltaic current was essential. And to prevent any idea that the chemical affinity was almost sufficient to decompose the water, and that a smaller current of electricity might, under the circumstances, cause the hydrogen to

* The acid was left during a night with a small piece of unamalgamated zinc in it, for the purpose of evolving such air as might be inclined to separate, and bringing the whole into a constant state.

† The experiment was repeated several times with the same results.

pass to the *cathode,* I need only refer to the results which I have given (807. 813.) to show that the chemical action at the electrodes has not the slightest influence over the *quantities* of water or other substances decomposed between them, but that they are entirely dependent upon the quantity of electricity which passes.

868. What, then, follows as a necessary consequence of the whole experiment? Why, this: that the chemical action upon 32·31 parts, or one equivalent of zinc, in this simple voltaic circle, was able to evolve such quantity of electricity in the form of a current as, passing through water, should decompose 9 parts, or one equivalent of that substance: and, considering the definite relations of electricity as developed in the preceding parts of the present paper, the results prove that the quantity of electricity which, being naturally associated with the particles of matter, gives them their combining power, is able, when thrown into a current, to separate those particles from their state of combination; or, in other words, that *the electricity which decomposes, and that which is evolved by the decomposition of, a certain quantity of matter, are alike.*

869. The harmony which this theory of the definite evolution and the equivalent definite action of electricity introduces into the associated theories of definite proportions and electro-chemical affinity, is very great. According to it, the equivalent weights of bodies are simply those quantities of them which contain equal quantities of electricity, or have naturally equal electric powers; it being the ELECTRICITY which *determines* the equivalent number, *because* it determines the combining force. Or, if we adopt the atomic theory or phraseology, then the atoms of bodies which are equivalents to each other in their ordinary chemical action, have equal quantities of electricity naturally associated with them. But I must confess I am jealous of the term *atom*; for though it is very easy to talk of atoms, it is very difficult to form a clear idea of their nature, especially when compound bodies are under consideration.

870. I cannot refrain from recalling here the beautiful idea put forth, I believe, by BERZELIUS (703.) in his development of his views of the electro-chemical theory of affinity, that the heat and light evolved during cases of powerful combination are the consequence of the electric discharge which is at the moment taking place. The idea is in perfect accordance with the view I have taken of the *quantity* of electricity associated with the particles of matter.

871. In this exposition of the law of the definite action of electricity, and its corresponding definite proportion in the particles of bodies, I do not pretend to have brought, as yet, every case of chemical or electro-chemical action under its dominion. There are numerous considerations of a theoretical nature, especially respecting the compound particles of matter and the resulting electrical forces which they ought to possess, which I hope will gradually receive their development; and there are numerous experimental cases, as, for instance, those of compounds formed by weak affinities, the simultaneous decomposition of water and salts, &c., which still require

investigation. But whatever the results on these and numerous other points may be, I do not believe that the facts which I have advanced, or even the general laws deduced from them, will suffer any serious change; and they are of sufficient importance to justify their publication, even though much may remain imperfect or undone. Indeed, it is the great beauty of our science, CHEMISTRY, that advancement in it, whether in a degree great or small, instead of exhausting the subjects of research, opens the doors to further and more abundant knowledge, overflowing with beauty and utility to those who will be at the easy personal pains of undertaking its experimental investigation.

872. The definite production of electricity (868.) in association with its definite action proves, I think, that the current of electricity in the voltaic pile is sustained by chemical decomposition, or rather by chemical action, and not by contact only. But here, as elsewhere (857.), I beg to reserve my opinion as to the real action of contact, not having yet been able to make up my mind as to its being either an exciting cause of the current, or merely necessary to allow of the conduction of electricity, otherwise generated, from one metal to the other.

873. But admitting that chemical action is the source of electricity, what an infinitely small fraction of that which is active do we obtain and employ in our voltaic batteries! Zinc and platina wires, one eighteenth of an inch in diameter and about half an inch long, dipped into dilute sulphuric acid, so weak that it is not sensibly sour to the tongue, or scarcely to our most delicate test papers, will evolve more electricity in one twentieth of a minute (860.) than any man would willingly allow to pass through his body at once. The chemical action of a grain of water upon four grains of zinc can evolve electricity equal in quantity to that of a powerful thunder-storm (868. 861.). Nor is it merely true that the quantity is active; it can be directed and made to perform its full equivalent duty (867. &c.). Is there not, then, great reason to hope and believe that, by a closer *experimental* investigation of the principles which govern the development and action of this subtile agent, we shall be able to increase the power of our batteries, or invent new instruments which shall a thousandfold surpass in energy those which we at present possess?

874. Here for a while I must leave the consideration of the *definite chemical action of electricity*. But before I dismiss this series of experimental Researches, I would call to mind that, in a former series, I showed the current of electricity was also *definite in its magnetic action* (366. 367. 376. 377.); and, though this result was not pursued to any extent, I have no doubt that the success which has attended the development of the chemical effects is not more than would accompany an investigation of the magnetic phenomena.

Royal Institution,
December 31*st*, 1833.

sliding vertical door containing a hole about $2\frac{1}{2}$ inches in diameter, which was accurately closed with a disk of coloured glass. In all cases the deposits were furthest from the light—and naturally so, seeing that a coloured object absorbs the heat more readily than a white one, and keeps the side of the bottle nearest to it of a higher temperature than the other parts.

One more point remains to be noticed. When mercury was exposed to the light in a tall narrow glass, no reliable results were obtained, that is, no deposit was formed that appeared to arise from the condensation of vapour. On two or three occasions metallic tears were seen in the vessel, but it was never clear to me that they did not arise from some shaking or disturbanee of the vessel. I could not reproduce even this unsatisfactory result in a narrow vessel, though I carefully tried for it by furnishing the vessel with a cap and stopcock and exhausting it with a syringe. I was also further surprised to find that a barometer-tube of thick glass charged with camphor and exhausted, produced little or no deposit even on the warmest days, and by exposure to direct sunshine. No sooner, however, had I dismissed the action of light from this subject, than the whole matter became clear. A thick glass tube by exposure to the light does not cool unequally, but slowly varies in temperature throughout its mass, so that no deposit either of mercury or of camphor is possible. If, however, the tube be thin, of large diameter and mounted, so that while one part is exposed to radiation the other part is protected, partial cooling is possible, and a deposit is produced. This, too, furnishes an explanation of a fact that had often surprised me. In barometers of large bore there is a deposit of mercury in the Torricellian vacuum on the side nearest the light. I had never seen this in a tube of small bore, though I had frequently looked for it in my own instrument. Some of the barometers of large bore in the International Exhibition have very fine deposits of mercury vapour in the Torricellian vacuum, but in such cases they are mounted so that the tube is more or less exposed. Where the tube is boxed in and protected from radiation there is little or no deposit.

King's College, London,
Long Vacation, 1862.

XLVIII. *Remarks on the Forces of Inorganic Nature.* *By* J. R. Mayer*.

THE following pages are designed as an attempt to answer the questions, What are we to understand by "Forces"?

* Translated from the *Annalen der Chemie und Pharmacie*, vol. xlii. p. 233 (May 1842), by G. C. Foster, B.A., Lecturer on Natural Philosophy

and how are different forces related to each other? Whereas the term *matter* implies the possession, by the object to which it is applied, of very definite properties, such as weight and extension; the term *force* conveys for the most part the idea of something unknown, unsearchable, and hypothetical. An attempt to render the notion of force equally exact with that of matter, and so to denote by it only objects of actual investigation, is one which, with the consequences that flow from it, ought not to be unwelcome to those who desire that their views of nature may be clear and unencumbered by hypotheses.

Forces are causes: accordingly, we may in relation to them make full application of the principle—*causa æquat effectum.* If the cause c has the effect e, then $c=e$; if, in its turn, e is the cause of a second effect f, we have $e=f$, and so on: $c=e=f\ldots=c$. In a chain of causes and effects, a term or a part of a term can never, as plainly appears from the nature of an equation, become equal to nothing. This first property of all causes we call their *indestructibility.*

If the given cause c has produced an effect e equal to itself, it has in that very act ceased to be: c has become e; if, after the production of e, c still remained in whole or in part, there must be still further effects corresponding to this remaining cause: the total effect of c would thus be $> e$, which would be contrary to the supposition $c=e$. Accordingly, since c becomes e, and e becomes f, &c., we must regard these various magnitudes as different forms under which one and the same object makes its appearance. This capability of assuming various forms is the second essential property of all causes. Taking both properties together, we may say, causes are (quantitatively) *indestructible* and (qualitatively) *convertible* objects.

Two classes of causes occur in nature, which, so far as experience goes, never pass one into another. The first class consists of such causes as possess the properties of weight and impenetrability; these are kinds of Matter: the other class is made up of causes which are wanting in the properties just mentioned, namely Forces, called also Imponderables, from the negative property that has been indicated. Forces are therefore *indestructible, convertible, imponderable objects.*

---

in Anderson's University, Glasgow.—Considerable attention having of late been called to the author of this paper, as one of the earliest propounders of the doctrine of the Indestructibility of Force, and especially of the idea of the equivalence of Heat and Work, it will probably interest many readers of the Philosophical Magazine to have placed in their hands his earliest publication on the subject. For some account of Mayer and of his further labours, see Prof. Tyndall's lecture "On Force," Phil. Mag. S. 4. vol. xxiv. pp. 64–66.

We will in the first instance take matter, to afford us an example of causes and effects. Explosive gas, H + O, and water, HO, are related to each other as cause and effect, therefore H + O = HO. But if H + O becomes HO, heat, *cal.*, makes its appearance as well as water; this heat must likewise have a cause, x, and we have therefore H + O + x = HO + *cal.* It might, however, be asked whether H + O is really = HO, and x = *cal.*, and not perhaps H + O = *cal.*, and x = HO, whence the above equation could equally be deduced; and so in many other cases. The phlogistic chemists recognized the equation between *cal.* and x, or Phlogiston as they called it, and in so doing made a great step in advance; but they involved themselves again in a system of mistakes by putting $-x$ in place of O; thus, for instance, they obtained H = HO + x.

Chemistry, whose problem it is to set forth in equations the causal connexion existing between the different kinds of matter, teaches us that matter, as a cause, has matter for its effect; but we are equally justified in saying that to force as cause, corresponds force as effect. Since $c=e$, and $e=c$, it is unnatural to call one term of an equation a force, and the other an effect of force or phenomenon, and to attach different notions to the expressions Force and Phenomenon. In brief, then, if the cause is matter, the effect is matter; if the cause is a force, the effect is also a force.

A cause which brings about the raising of a weight is a force; its effect (*the raised weight*) is, accordingly, equally *a force*; or, expressing this relation in a more general form, *separation in space of ponderable objects is a force*; since this force causes the fall of bodies, we call it *falling force*. Falling force and fall, or, more generally still, falling force and motion, are forces which are related to each other as cause and effect—forces which are convertible one into the other—two different forms of one and the same object. For example, a weight resting on the ground is not a force: it is neither the cause of motion, nor of the lifting of another weight; it becomes so, however, in proportion as it is raised above the ground: the cause—the distance between a weight and the earth—and the effect—the quantity of motion produced—bear to each other, as we learn from mechanics, a constant relation.

Gravity being regarded as the cause of the falling of bodies, a gravitating force is spoken of, and so the notions of *property* and of *force* are confounded with each other: precisely that which is the essential attribute of every force—the *union* of indestructibility with convertibility—is wanting in every property: between a property and a force, between gravity and motion, it is therefore impossible to establish the equation required for a rightly conceived causal relation. If gravity be called a force, a cause is supposed which produces effects without itself diminishing, and incorrect conceptions of the causal connexion of things are thereby fostered. In order that a body may fall, it is no less necessary that it should be lifted up, than that it should be heavy or possess gravity; the fall of bodies ought not therefore to be ascribed to their gravity alone.

It is the problem of Mechanics to develope the equations which subsist between falling force and motion, motion and falling force, and between different motions: here we will call to mind only one point. The magnitude of the falling force v is directly proportional (the earth's radius being assumed $=\infty$) to the magnitude of the mass m, and the height d to which it is raised; that is, $v=md$. If the height $d=1$, to which the mass m is raised, is transformed into the final velocity $c=1$ of this mass, we have also $v=mc$; but from the known relations existing between d and c, it results that, for other values of d or of c, the measure of the force v is mc^2; accordingly $v=md=mc^2$: the law of the conservation of *vis viva* is thus found to be based on the general law of the indestructibility of causes.

In numberless cases we see motion cease without having caused another motion or the lifting of a weight; but a force once in existence cannot be annihilated, it can only change its form; and the question therefore arises, What other forms is force, which we have become acquainted with as falling force and motion, capable of assuming? Experience alone can lead us to a conclusion on this point. In order to experiment with advantage, we must select implements which, besides causing a real cessation of motion, are as little as possible altered by the objects to be examined. If, for example, we rub together two metal plates, we see motion disappear, and heat, on the other hand, make its appearance, and we have now only to ask whether *motion* is the cause of heat. In order to come to a decision on this point, we must discuss the question whether, in the numberless cases in which the expenditure of motion is accompanied by the appearance of heat, the motion has not some other effect than the production of heat, and the heat some other cause than the motion.

An attempt to ascertain the effects of ceasing motion has never yet been seriously made; without, therefore, wishing to exclude *à priori* the hypotheses which it may be possible to set up, we observe only that, as a rule, this effect cannot be supposed to be an alteration in the state of aggregation of the moved (that is, rubbing, &c.) bodies. If we assume that a certain quantity of motion v is expended in the conversion of a rubbing substance m into n, we must then have $m+v=n$, and $n=m+v$; and when n is reconverted into m, v must appear again in some form or other. By the friction of two metallic plates continued for a very long time, we can gradually cause the cessation of an immense quantity of movement; but would it ever occur to us to look for even the smallest trace of the force which has disappeared in the metallic dust that we could collect, and to try to regain it thence? We repeat, the motion cannot have been annihilated; and contrary, or positive and negative, motions cannot be regarded as =0, any more than contrary motions can come out of nothing, or a weight can raise itself.

Without the recognition of a causal connexion between motion and heat, it is just as difficult to explain the production of heat as it is to give any account of the motion that disappears. The heat cannot be derived from the diminution of the volume of the rubbing substances. It is well known that two pieces of ice may be melted by rubbing them together *in vacuo*; but let any one try to convert ice into water by pressure*, however enormous. Water undergoes, as was found by the author, a rise of temperature when violently shaken. The water so heated (from 12° to 13° C.) has a greater bulk after being shaken than it had before; whence now comes this quantity of heat, which by repeated shaking may be called into existence in the same apparatus as often as we please? The vibratory hypothesis of heat is an approach towards the doctrine of heat being the effect of motion, but it does not favour the admission of this causal relation in its full generality; it rather lays the chief stress on uneasy oscillations (*unbehagliche Schwingungen*).

If it be now considered as established that in many cases (*exceptio confirmat regulam*) no other effect of motion can be traced except heat, and that no other cause than motion can be found for the heat that is produced, we prefer the assumption that heat proceeds from motion, to the assumption of a cause without effect and of an effect without a cause,—just as the chemist, instead of allowing oxygen and hydrogen to disappear without further investigation, and water to be produced in some inexplicable manner, establishes a connexion between oxygen and hydrogen on the one hand and water on the other.

The natural connexion existing between falling force, motion, and heat may be conceived of as follows. We know that heat makes its appearance when the separate particles of a body approach nearer to each other: condensation produces heat.

* Since the original publication of this paper, Prof. W. Thomson has shown that pressure has a sensible effect in liquefying ice (*Conf.* Phil. Mag. S. 3. vol. xxxvii. p. 123); but the experiments of Bunsen and of Hopkins have shown that the melting-points of bodies which expand on becoming liquid are raised by pressure, which is all that Mayer's argument requires.—G. C. F.

And what applies to the smallest particles of matter, and the smallest intervals between them, must also apply to large masses and to measureable distances. The falling of a weight is a real diminution of the bulk of the earth, and must therefore without doubt be related to the quantity of heat thereby developed; this quantity of heat must be proportional to the greatness of the weight and its distance from the ground. From this point of view we are very easily led to the equations between falling force, motion, and heat, that have already been discussed.

But just as little as the connexion between falling force and motion authorizes the conclusion that the essence of falling force is motion, can such a conclusion be adopted in the case of heat. We are, on the contrary, rather inclined to infer that, before it can become heat, motion—whether simple, or vibratory as in the case of light and radiant heat, &c.—must cease to exist as motion.

If falling force and motion are equivalent to heat, heat must also naturally be equivalent to motion and falling force. Just as heat appears as an *effect* of the diminution of bulk and of the cessation of motion, so also does heat disappear as a *cause* when its effects are produced in the shape of motion, expansion, or raising of weight.

In water-mills, the continual diminution in bulk which the earth undergoes, owing to the fall of the water, gives rise to motion, which afterwards disappears again, calling forth unceasingly a great quantity of heat; and inversely, the steam-engine serves to decompose heat again into motion or the raising of weights. A locomotive engine with its train may be compared to a distilling apparatus; the heat applied under the boiler passes off as motion, and this is deposited again as heat at the axles of the wheels.

We will close our disquisition, the propositions of which have resulted as necessary consequences from the principle "causa æquat effectum," and which are in accordance with all the phenomena of Nature, with a practical deduction. The solution of the equations subsisting between falling force and motion requires that the space fallen through in a given time, *e. g.* the first second, should be experimentally determined; in like manner, the solution of the equations subsisting between falling force and motion on the one hand and heat on the other, requires an answer to the question, How great is the quantity of heat which corresponds to a given quantity of motion or falling force? For instance, we must ascertain how high a given weight requires to be raised above the ground in order that its falling force may be equivalent to the raising of the temperature of an equal weight of water from 0° to 1° C. The attempt to show that such an equation is the expression of a physical truth may be regarded as the substance of the foregoing remarks.

By applying the principles that have been set forth to the relations subsisting between the temperature and the volume of gases, we find that the sinking of a mercury column by which a gas is compressed is equivalent to the quantity of heat set free by the compression; and hence it follows, the ratio between the capacity for heat of air under constant pressure and its capacity under constant volume being taken as $=1\cdot421$, that the warming of a given weight of water from 0° to 1° C. corresponds to the fall of an equal weight from the height of about 365 metres*. If we compare with this result the working of our best steam-engines, we see how small a part only of the heat applied under the boiler is really transformed into motion or the raising of weights; and this may serve as justification for the attempts at the profitable production of motion by some other method than the expenditure of the chemical difference between carbon and oxygen—more particularly by the transformation into motion of electricity obtained by chemical means.

XLIX. *The Excavation of the Valleys of the Alps.* *By* A. C. Ramsay, *F.R.S.*†

IN the month of March last I read a memoir to the Geological Society on the Glacial Origin of the Swiss and other Lakes, which has since been published in that Society's Quarterly Journal for August. In that memoir I incidentally alluded (p. 200) to the existence of the chief Alpine valleys *before* the glaciers attained their greatest extension, which valleys were afterwards "*modified* in form by the weight and grinding power of ice in motion."

In a previous memoir, published in 1859, I stated that "it is certain all glaciers must deepen their beds by erosion, and it may be that, when a glacier filled a valley" almost to the brim, "the thickness of the ice was not equal to the present mass added to the superincumbent weight indicated by the signs (striation, &c.) on the slopes above the present surface of the glacier." But though glaciers certainly have a powerful effect in deepening their beds, it has always appeared to me a difficult and perhaps an impossible point to determine to what extent the great Alpine valleys have been eroded by ice—whether, in fact, they have been chiefly scooped out by it, or whether, as I always believed,

* When the corrected specific heat of air is introduced into the calculation this number is increased, and agrees then with the experimental determinations of Mr. Joule.

† Communicated by the Author.

VIII. *A Dynamical Theory of the Electromagnetic Field.* *By* J. Clerk Maxwell, *F.R.S.*

Received October 27,—Read December 8, 1864.

PART I.—INTRODUCTORY.

(1) The most obvious mechanical phenomenon in electrical and magnetical experiments is the mutual action by which bodies in certain states set each other in motion while still at a sensible distance from each other. The first step, therefore, in reducing these phenomena into scientific form, is to ascertain the magnitude and direction of the force acting between the bodies, and when it is found that this force depends in a certain way upon the relative position of the bodies and on their electric or magnetic condition, it seems at first sight natural to explain the facts by assuming the existence of something either at rest or in motion in each body, constituting its electric or magnetic state, and capable of acting at a distance according to mathematical laws.

In this way mathematical theories of statical electricity, of magnetism, of the mechanical action between conductors carrying currents, and of the induction of currents have been formed. In these theories the force acting between the two bodies is treated with reference only to the condition of the bodies and their relative position, and without any express consideration of the surrounding medium.

These theories assume, more or less explicitly, the existence of substances the particles of which have the property of acting on one another at a distance by attraction or repulsion. The most complete development of a theory of this kind is that of M. W. Weber*, who has made the same theory include electrostatic and electromagnetic phenomena.

In doing so, however, he has found it necessary to assume that the force between two electric particles depends on their relative velocity, as well as on their distance.

This theory, as developed by MM. W. Weber and C. Neumann†, is exceedingly ingenious, and wonderfully comprehensive in its application to the phenomena of statical electricity, electromagnetic attractions, induction of currents and diamagnetic phenomena; and it comes to us with the more authority, as it has served to guide the speculations of one who has made so great an advance in the practical part of electric science, both by introducing a consistent system of units in electrical measurement, and by actually determining electrical quantities with an accuracy hitherto unknown.

* Electrodynamische Maassbestimmungen. Leipzic Trans. vol. i. 1849, and Taylor's Scientific Memoirs, vol. v. art. xiv.

† "Explicare tentatur quomodo fiat ut lucis planum polarizationis per vires electricas vel magneticas declinetur."—Halis Saxonum, 1858.

(2) The mechanical difficulties, however, which are involved in the assumption of particles acting at a distance with forces which depend on their velocities are such as to prevent me from considering this theory as an ultimate one, though it may have been, and may yet be useful in leading to the coordination of phenomena.

I have therefore preferred to seek an explanation of the fact in another direction, by supposing them to be produced by actions which go on in the surrounding medium as well as in the excited bodies, and endeavouring to explain the action between distant bodies without assuming the existence of forces capable of acting directly at sensible distances.

(3) The theory I propose may therefore be called a theory of the *Electromagnetic Field*, because it has to do with the space in the neighbourhood of the electric or magnetic bodies, and it may be called a *Dynamical* Theory, because it assumes that in that space there is matter in motion, by which the observed electromagnetic phenomena are produced.

(4) The electromagnetic field is that part of space which contains and surrounds bodies in electric or magnetic conditions.

It may be filled with any kind of matter, or we may endeavour to render it empty of all gross matter, as in the case of GEISSLER's tubes and other so-called vacua.

There is always, however, enough of matter left to receive and transmit the undulations of light and heat, and it is because the transmission of these radiations is not greatly altered when transparent bodies of measurable density are substituted for the so-called vacuum, that we are obliged to admit that the undulations are those of an æthereal substance, and not of the gross matter, the presence of which merely modifies in some way the motion of the æther.

We have therefore some reason to believe, from the phenomena of light and heat, that there is an æthereal medium filling space and permeating bodies, capable of being set in motion and of transmitting that motion from one part to another, and of communicating that motion to gross matter so as to heat it and affect it in various ways.

(5) Now the energy communicated to the body in heating it must have formerly existed in the moving medium, for the undulations had left the source of heat some time before they reached the body, and during that time the energy must have been half in the form of motion of the medium and half in the form of elastic resilience. From these considerations Professor W. THOMSON has argued*, that the medium must have a density capable of comparison with that of gross matter, and has even assigned an inferior limit to that density.

(6) We may therefore receive, as a datum derived from a branch of science independent of that with which we have to deal, the existence of a pervading medium, of small but real density, capable of being set in motion, and of transmitting motion from one part to another with great, but not infinite, velocity.

Hence the parts of this medium must be so connected that the motion of one part

* "On the Possible Density of the Luminiferous Medium, and on the Mechanical Value of a Cubic Mile of Sunlight," Transactions of the Royal Society of Edinburgh (1854), p. 57.

depends in some way on the motion of the rest; and at the same time these connexions must be capable of a certain kind of elastic yielding, since the communication of motion is not instantaneous, but occupies time.

The medium is therefore capable of receiving and storing up two kinds of energy, namely, the "actual" energy depending on the motions of its parts, and "potential" energy, consisting of the work which the medium will do in recovering from displacement in virtue of its elasticity.

The propagation of undulations consists in the continual transformation of one of these forms of energy into the other alternately, and at any instant the amount of energy in the whole medium is equally divided, so that half is energy of motion, and half is elastic resilience.

(7) A medium having such a constitution may be capable of other kinds of motion and displacement than those which produce the phenomena of light and heat, and some of these may be of such a kind that they may be evidenced to our senses by the phenomena they produce.

(8) Now we know that the luminiferous medium is in certain cases acted on by magnetism; for FARADAY* discovered that when a plane polarized ray traverses a transparent diamagnetic medium in the direction of the lines of magnetic force produced by magnets or currents in the neighbourhood, the plane of polarization is caused to rotate.

This rotation is always in the direction in which positive electricity must be carried round the diamagnetic body in order to produce the actual magnetization of the field.

M. VERDET† has since discovered that if a paramagnetic body, such as solution of perchloride of iron in ether, be substituted for the diamagnetic body, the rotation is in the opposite direction.

Now Professor W. THOMSON‡ has pointed out that no distribution of forces acting between the parts of a medium whose only motion is that of the luminous vibrations, is sufficient to account for the phenomena, but that we must admit the existence of a motion in the medium depending on the magnetization, in addition to the vibratory motion which constitutes light.

It is true that the rotation by magnetism of the plane of polarization has been observed only in media of considerable density; but the properties of the magnetic field are not so much altered by the substitution of one medium for another, or for a vacuum, as to allow us to suppose that the dense medium does anything more than merely modify the motion of the ether. We have therefore warrantable grounds for inquiring whether there may not be a motion of the ethereal medium going on wherever magnetic effects are observed, and we have some reason to suppose that this motion is one of rotation, having the direction of the magnetic force as its axis.

(9) We may now consider another phenomenon observed in the electromagnetic

* Experimental Researches, Series 19.

† Comptes Rendus (1856, second half year, p. 529, and 1857, first half year, p. 1209).

‡ Proceedings of the Royal Society, June 1856 and June 1861.

field. When a body is moved across the lines of magnetic force it experiences what is called an electromotive force; the two extremities of the body tend to become oppositely electrified, and an electric current tends to flow through the body. When the electromotive force is sufficiently powerful, and is made to act on certain compound bodies, it decomposes them, and causes one of their components to pass towards one extremity of the body, and the other in the opposite direction.

Here we have evidence of a force causing an electric current in spite of resistance; electrifying the extremities of a body in opposite ways, a condition which is sustained only by the action of the electromotive force, and which, as soon as that force is removed, tends, with an equal and opposite force, to produce a counter current through the body and to restore the original electrical state of the body; and finally, if strong enough, tearing to pieces chemical compounds and carrying their components in opposite directions, while their natural tendency is to combine, and to combine with a force which can generate an electromotive force in the reverse direction.

This, then, is a force acting on a body caused by its motion through the electromagnetic field, or by changes occurring in that field itself; and the effect of the force is either to produce a current and heat the body, or to decompose the body, or, when it can do neither, to put the body in a state of electric polarization,—a state of constraint in which opposite extremities are oppositely electrified, and from which the body tends to relieve itself as soon as the disturbing force is removed.

(10) According to the theory which I propose to explain, this "electromotive force" is the force called into play during the communication of motion from one part of the medium to another, and it is by means of this force that the motion of one part causes motion in another part. When electromotive force acts on a conducting circuit, it produces a current, which, as it meets with resistance, occasions a continual transformation of electrical energy into heat, which is incapable of being restored again to the form of electrical energy by any reversal of the process.

(11) But when electromotive force acts on a dielectric it produces a state of polarization of its parts similar in distribution to the polarity of the parts of a mass of iron under the influence of a magnet, and like the magnetic polarization, capable of being described as a state in which every particle has its opposite poles in opposite conditions*.

In a dielectric under the action of electromotive force, we may conceive that the electricity in each molecule is so displaced that one side is rendered positively and the other negatively electrical, but that the electricity remains entirely connected with the molecule, and does not pass from one molecule to another. The effect of this action on the whole dielectric mass is to produce a general displacement of electricity in a certain direction. This displacement does not amount to a current, because when it has attained to a certain value it remains constant, but it is the commencement of a current, and its variations constitute currents in the positive or the negative direction according

* FARADAY, Exp. Res. Series XI.; MOSSOTTI, Mem. della Soc. Italiana (Modena), vol. xxiv. part 2. p. 49.

as the displacement is increasing or decreasing. In the interior of the dielectric there is no indication of electrification, because the electrification of the surface of any molecule is neutralized by the opposite electrification of the surface of the molecules in contact with it; but at the bounding surface of the dielectric, where the electrification is not neutralized, we find the phenomena which indicate positive or negative electrification.

The relation between the electromotive force and the amount of electric displacement it produces depends on the nature of the dielectric, the same electromotive force producing generally a greater electric displacement in solid dielectrics, such as glass or sulphur, than in air.

(12) Here, then, we perceive another effect of electromotive force, namely, electric displacement, which according to our theory is a kind of elastic yielding to the action of the force, similar to that which takes place in structures and machines owing to the want of perfect rigidity of the connexions.

(13) The practical investigation of the inductive capacity of dielectrics is rendered difficult on account of two disturbing phenomena. The first is the conductivity of the dielectric, which, though in many cases exceedingly small, is not altogether insensible. The second is the phenomenon called electric absorption*, in virtue of which, when the dielectric is exposed to electromotive force, the electric displacement gradually increases, and when the electromotive force is removed, the dielectric does not instantly return to its primitive state, but only discharges a portion of its electrification, and when left to itself gradually acquires electrification on its surface, as the interior gradually becomes depolarized. Almost all solid dielectrics exhibit this phenomenon, which gives rise to the residual charge in the Leyden jar, and to several phenomena of electric cables described by Mr. F. JENKIN†.

(14) We have here two other kinds of yielding besides the yielding of the perfect dielectric, which we have compared to a perfectly elastic body. The yielding due to conductivity may be compared to that of a viscous fluid (that is to say, a fluid having great internal friction), or a soft solid on which the smallest force produces a permanent alteration of figure increasing with the time during which the force acts. The yielding due to electric absorption may be compared to that of a cellular elastic body containing a thick fluid in its cavities. Such a body, when subjected to pressure, is compressed by degrees on account of the gradual yielding of the thick fluid; and when the pressure is removed it does not at once recover its figure, because the elasticity of the substance of the body has gradually to overcome the tenacity of the fluid before it can regain complete equilibrium.

Several solid bodies in which no such structure as we have supposed can be found, seem to possess a mechanical property of this kind‡; and it seems probable that the

* FARADAY, Exp. Res. 1233–1250.

† Reports of British Association, 1859, p. 248; and Report of Committee of Board of Trade on Submarine Cables, pp. 136 & 464.

‡ As, for instance, the composition of glue, treacle, &c., of which small plastic figures are made, which after being distorted gradually recover their shape.

same substances, if dielectrics, may possess the analogous electrical property, and if magnetic, may have corresponding properties relating to the acquisition, retention, and loss of magnetic polarity.

(15) It appears therefore that certain phenomena in electricity and magnetism lead to the same conclusion as those of optics, namely, that there is an æthereal medium pervading all bodies, and modified only in degree by their presence; that the parts of this medium are capable of being set in motion by electric currents and magnets; that this motion is communicated from one part of the medium to another by forces arising from the connexions of those parts; that under the action of these forces there is a certain yielding depending on the elasticity of these connexions; and that therefore energy in two different forms may exist in the medium, the one form being the actual energy of motion of its parts, and the other being the potential energy stored up in the connexions, in virtue of their elasticity.

(16) Thus, then, we are led to the conception of a complicated mechanism capable of a vast variety of motion, but at the same time so connected that the motion of one part depends, according to definite relations, on the motion of other parts, these motions being communicated by forces arising from the relative displacement of the connected parts, in virtue of their elasticity. Such a mechanism must be subject to the general laws of Dynamics, and we ought to be able to work out all the consequences of its motion, provided we know the form of the relation between the motions of the parts.

(17) We know that when an electric current is established in a conducting circuit, the neighbouring part of the field is characterized by certain magnetic properties, and that if two circuits are in the field, the magnetic properties of the field due to the two currents are combined. Thus each part of the field is in connexion with both currents, and the two currents are put in connexion with each other in virtue of their connexion with the magnetization of the field. The first result of this connexion that I propose to examine, is the induction of one current by another, and by the motion of conductors in the field.

The second result, which is deduced from this, is the mechanical action between conductors carrying currents. The phenomenon of the induction of currents has been deduced from their mechanical action by HELMHOLTZ * and THOMSON †. I have followed the reverse order, and deduced the mechanical action from the laws of induction. I have then described experimental methods of determining the quantities L, M, N, on which these phenomena depend.

(18) I then apply the phenomena of induction and attraction of currents to the exploration of the electromagnetic field, and the laying down systems of lines of magnetic force which indicate its magnetic properties. By exploring the same field with a magnet, I show the distribution of its equipotential magnetic surfaces, cutting the lines of force at right angles.

* "Conservation of Force," Physical Society of Berlin, 1847; and TAYLOR's Scientific Memoirs, 1853, p. 114.

† Reports of the British Association, 1848; Philosophical Magazine, Dec. 1851.

In order to bring these results within the power of symbolical calculation, I then express them in the form of the General Equations of the Electromagnetic Field. These equations express—

(A) The relation between electric displacement, true conduction, and the total current, compounded of both.

(B) The relation between the lines of magnetic force and the inductive coefficients of a circuit, as already deduced from the laws of induction.

(C) The relation between the strength of a current and its magnetic effects, according to the electromagnetic system of measurement.

(D) The value of the electromotive force in a body, as arising from the motion of the body in the field, the alteration of the field itself, and the variation of electric potential from one part of the field to another.

(E) The relation between electric displacement, and the electromotive force which produces it.

(F) The relation between an electric current, and the electromotive force which produces it.

(G) The relation between the amount of free electricity at any point, and the electric displacements in the neighbourhood.

(H) The relation between the increase or diminution of free electricity and the electric currents in the neighbourhood.

There are twenty of these equations in all, involving twenty variable quantities.

(19) I then express in terms of these quantities the intrinsic energy of the Electromagnetic Field as depending partly on its magnetic and partly on its electric polarization at every point.

From this I determine the mechanical force acting, 1st, on a moveable conductor carrying an electric current; 2ndly, on a magnetic pole; 3rdly, on an electrified body.

The last result, namely, the mechanical force acting on an electrified body, gives rise to an independent method of electrical measurement founded on its electrostatic effects. The relation between the units employed in the two methods is shown to depend on what I have called the "electric elasticity" of the medium, and to be a velocity, which has been experimentally determined by MM. WEBER and KOHLRAUSCH.

I then show how to calculate the electrostatic capacity of a condenser, and the specific inductive capacity of a dielectric.

The case of a condenser composed of parallel layers of substances of different electric resistances and inductive capacities is next examined, and it is shown that the phenomenon called electric absorption will generally occur, that is, the condenser, when suddenly discharged, will after a short time show signs of a *residual* charge.

(20) The general equations are next applied to the case of a magnetic disturbance propagated through a non-conducting field, and it is shown that the only disturbances which can be so propagated are those which are transverse to the direction of propagation, and that the velocity of propagation is the velocity v, found from experiments such

as those of WEBER, which expresses the number of electrostatic units of electricity which are contained in one electromagnetic unit.

This velocity is so nearly that of light, that it seems we have strong reason to conclude that light itself (including radiant heat, and other radiations if any) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws. If so, the agreement between the elasticity of the medium as calculated from the rapid alternations of luminous vibrations, and as found by the slow processes of electrical experiments, shows how perfect and regular the elastic properties of the medium must be when not encumbered with any matter denser than air. If the same character of the elasticity is retained in dense transparent bodies, it appears that the square of the index of refraction is equal to the product of the specific dielectric capacity and the specific magnetic capacity. Conducting media are shown to absorb such radiations rapidly, and therefore to be generally opaque.

The conception of the propagation of transverse magnetic disturbances to the exclusion of normal ones is distinctly set forth by Professor FARADAY* in his "Thoughts on Ray Vibrations." The electromagnetic theory of light, as proposed by him, is the same in substance as that which I have begun to develope in this paper, except that in 1846 there were no data to calculate the velocity of propagation.

(21) The general equations are then applied to the calculation of the coefficients of mutual induction of two circular currents and the coefficient of self-induction in a coil. The want of uniformity of the current in the different parts of the section of a wire at the commencement of the current is investigated, I believe for the first time, and the consequent correction of the coefficient of self-induction is found.

These results are applied to the calculation of the self-induction of the coil used in the experiments of the Committee of the British Association on Standards of Electric Resistance, and the value compared with that deduced from the experiments.

PART II.—ON ELECTROMAGNETIC INDUCTION.

Electromagnetic Momentum of a Current.

(22) We may begin by considering the state of the field in the neighbourhood of an electric current. We know that magnetic forces are excited in the field, their direction and magnitude depending according to known laws upon the form of the conductor carrying the current. When the strength of the current is increased, all the magnetic effects are increased in the same proportion. Now, if the magnetic state of the field depends on motions of the medium, a certain force must be exerted in order to increase or diminish these motions, and when the motions are excited they continue, so that the effect of the connexion between the current and the electromagnetic field surrounding it, is to endow the current with a kind of momentum, just as the connexion between the driving-point of a machine and a fly-wheel endows the driving-point with an addi-

* Philosophical Magazine, May 1846, or Experimental Researches, iii. p. 447.

tional momentum, which may be called the momentum of the fly-wheel reduced to the driving-point. The unbalanced force acting on the driving-point increases this momentum, and is measured by the rate of its increase.

In the case of electric currents, the resistance to sudden increase or diminution of strength produces effects exactly like those of momentum, but the amount of this momentum depends on the shape of the conductor and the relative position of its different parts.

Mutual Action of two Currents.

(23) If there are two electric currents in the field, the magnetic force at any point is that compounded of the forces due to each current separately, and since the two currents are in connexion with every point of the field, they will be in connexion with each other, so that any increase or diminution of the one will produce a force acting with or contrary to the other.

Dynamical Illustration of Reduced Momentum.

(24) As a dynamical illustration, let us suppose a body C so connected with two independent driving-points A and B that its velocity is p times that of A together with q times that of B. Let u be the velocity of A, v that of B, and w that of C, and let δx, δy, δz be their simultaneous displacements, then by the general equation of dynamics*,

$$C\frac{dw}{dt}\delta z=X\delta x+Y\delta y,$$

where X and Y are the forces acting at A and B.

But

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

and

$$\delta z=p\delta x+q\delta y.$$

Substituting, and remembering that δx and δy are independent,

$$\left.\begin{aligned}X&=\frac{d}{dt}(Cp^2u+Cpqv),\\ Y&=\frac{d}{dt}(Cpqu+Cq^2v).\end{aligned}\right\} \quad \ldots \ldots \ldots \quad (1)$$

We may call $Cp^2u+Cpqv$ the momentum of C referred to A, and $Cpqu+Cq^2v$ its momentum referred to B; then we may say that the effect of the force X is to increase the momentum of C referred to A, and that of Y to increase its momentum referred to B.

If there are many bodies connected with A and B in a similar way but with different values of p and q, we may treat the question in the same way by assuming

$$L=\Sigma(Cp^2), \quad M=\Sigma(Cpq), \quad \text{and } N=\Sigma(Cq^2),$$

* LAGRANGE, Méc. Anal. ii. 2. § 5.

where the summation is extended to all the bodies with their proper values of C, p, and q. Then the momentum of the system referred to A is

$$\mathrm{L}u+\mathrm{M}v,$$

and referred to B,

$$\mathrm{M}u+\mathrm{N}v,$$

and we shall have

$$\left.\begin{aligned} \mathrm{X}&=\frac{d}{dt}(\mathrm{L}u+\mathrm{M}v), \\ \mathrm{Y}&=\frac{d}{dt}(\mathrm{M}u+\mathrm{N}v), \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where X and Y are the external forces acting on A and B.

(25) To make the illustration more complete we have only to suppose that the motion of A is resisted by a force proportional to its velocity, which we may call Ru, and that of B by a similar force, which we may call Sv, R and S being coefficients of resistance. Then if ξ and η are the forces on A and B

$$\left.\begin{aligned} \xi&=\mathrm{X}+\mathrm{R}u=\mathrm{R}u+\frac{d}{dt}(\mathrm{L}u+\mathrm{M}v), \\ \eta&=\mathrm{Y}+\mathrm{S}v=\mathrm{S}v+\frac{d}{dt}(\mathrm{M}u+\mathrm{N}v) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If the velocity of A be increased at the rate $\frac{du}{dt}$, then in order to prevent B from moving a force, $\eta=\frac{d}{dt}(\mathrm{M}u)$ must be applied to it.

This effect on B, due to an increase of the velocity of A, corresponds to the electromotive force on one circuit arising from an increase in the strength of a neighbouring circuit.

This dynamical illustration is to be considered merely as assisting the reader to understand what is meant in mechanics by Reduced Momentum. The facts of the induction of currents as depending on the variations of the quantity called Electromagnetic Momentum, or Electrotonic State, rest on the experiments of FARADAY*, FELICI†, &c.

Coefficients of Induction for Two Circuits.

(26) In the electromagnetic field the values of L, M, N depend on the distribution of the magnetic effects due to the two circuits, and this distribution depends only on the form and relative position of the circuits. Hence L, M, N are quantities depending on the form and relative position of the circuits, and are subject to variation with the motion of the conductors. It will be presently seen that L, M, N are geometrical quantities of the nature of lines, that is, of one dimension in space; L depends on the form of the first conductor, which we shall call A, N on that of the second, which we shall call B, and M on the relative position of A and B.

(27) Let ξ be the electromotive force acting on A, x the strength of the current, and

* Experimental Researches, Series I., IX.

† Annales de Chimie, sér. 3. xxxiv. (1852) p. 64.

R the resistance, then Rx will be the resisting force. In steady currents the electromotive force just balances the resisting force, but in variable currents the resultant force $\xi = \mathrm{R}x$ is expended in increasing the "electromagnetic momentum," using the word momentum merely to express that which is generated by a force acting during a time, that is, a velocity existing in a body.

In the case of electric currents, the force in action is not ordinary mechanical force, at least we are not as yet able to measure it as common force, but we call it electromotive force, and the body moved is not merely the electricity in the conductor, but something outside the conductor, and capable of being affected by other conductors in the neighbourhood carrying currents. In this it resembles rather the reduced momentum of a driving-point of a machine as influenced by its mechanical connexions, than that of a simple moving body like a cannon ball, or water in a tube.

Electromagnetic Relations of two Conducting Circuits.

(28.) In the case of two conducting circuits, A and B, we shall assume that the electromagnetic momentum belonging to A is

$$\mathrm{L}x + \mathrm{M}y,$$

and that belonging to B,

$$\mathrm{M}x + \mathrm{N}y,$$

where L, M, N correspond to the same quantities in the dynamical illustration, except that they are supposed to be capable of variation when the conductors A or B are moved.

Then the equation of the current x in A will be

$$\xi = \mathrm{R}x + \frac{d}{dt}(\mathrm{L}x + \mathrm{M}y), \quad \ldots \quad (4)$$

and that of y in B

$$\eta = \mathrm{S}y + \frac{d}{dt}(\mathrm{M}x + \mathrm{N}y), \quad \ldots \quad (5)$$

where ξ and η are the electromotive forces, x and y the currents, and R and S the resistances in A and B respectively.

Induction of one Current by another.

(29) Case 1st. Let there be no electromotive force on B, except that which arises from the action of A, and let the current of A increase from 0 to the value x, then

$$\mathrm{S}y + \frac{d}{dt}(\mathrm{M}x + \mathrm{N}y) = 0,$$

whence

$$\mathrm{Y} = \int_0^t y\,dt = -\frac{\mathrm{M}}{\mathrm{S}}x,$$

that is, a quantity of electricity Y, being the total induced current, will flow through B when x rises from 0 to x. This is induction by variation of the current in the primary conductor. When M is positive, the induced current due to increase of the primary current is negative.

Induction by Motion of Conductor.

(30) Case 2nd. Let x remain constant, and let M change from M to M′, then

$$Y = -\frac{M'-M}{S}x;$$

so that if M is increased, which it will be by the primary and secondary circuits approaching each other, there will be a negative induced current, the total quantity of electricity passed through B being Y.

This is induction by the relative motion of the primary and secondary conductors.

Equation of Work and Energy.

(31) To form the equation between work done and energy produced, multiply (1) by x and (2) by y, and add

$$\xi x + \eta y = Rx^2 + Sy^2 + x\frac{d}{dt}(Lx+My) + y\frac{d}{dt}(Mx+Ny). \quad . \quad . \quad . \quad . \quad (8)$$

Here ξx is the work done in unit of time by the electromotive force ξ acting on the current x and maintaining it, and ηy is the work done by the electromotive force η. Hence the left-hand side of the equation represents the work done by the electromotive forces in unit of time.

Heat produced by the Current.

(32) On the other side of the equation we have, first,

$$Rx^2 + Sy^2 = H, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

which represents the work done in overcoming the resistance of the circuits in unit of time. This is converted into heat. The remaining terms represent work not converted into heat. They may be written

$$\tfrac{1}{2}\frac{d}{dt}(Lx^2+2Mxy+Ny^2) + \tfrac{1}{2}\frac{dL}{dt}x^2 + \frac{dM}{dt}xy + \tfrac{1}{2}\frac{dN}{dt}y^2.$$

Intrinsic Energy of the Currents.

(33) If L, M, N are constant, the whole work of the electromotive forces which is not spent against resistance will be devoted to the development of the currents. The whole intrinsic energy of the currents is therefore

$$\tfrac{1}{2}Lx^2 + Mxy + \tfrac{1}{2}Ny^2 = E. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

This energy exists in a form imperceptible to our senses, probably as actual motion, the seat of this motion being not merely the conducting circuits, but the space surrounding them.

Mechanical Action between Conductors.

(34) The remaining terms,

$$\tfrac{1}{2}\frac{d\mathrm{L}}{dt}x^2+\frac{d\mathrm{M}}{dt}xy+\tfrac{1}{2}\frac{d\mathrm{N}}{dt}y^2=\mathrm{W} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

represent the work done in unit of time arising from the variations of L, M, and N, or, what is the same thing, alterations in the form and position of the conducting circuits A and B.

Now if work is done when a body is moved, it must arise from ordinary mechanical force acting on the body while it is moved. Hence this part of the expression shows that there is a mechanical force urging every part of the conductors themselves in that direction in which L, M, and N will be most increased.

The existence of the electromagnetic force between conductors carrying currents is therefore a direct consequence of the joint and independent action of each current on the electromagnetic field. If A and B are allowed to approach a distance ds, so as to increase M from M to M′ while the currents are x and y, then the work done will be

$$(\mathrm{M}'-\mathrm{M})xy,$$

and the force in the direction of ds will be

$$\frac{d\mathrm{M}}{ds}xy, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

and this will be an attraction if x and y are of the same sign, and if M is increased as A and B approach.

It appears, therefore, that if we admit that the unresisted part of electromotive force goes on as long as it acts, generating a self-persistent state of the current, which we may call (from mechanical analogy) its electromagnetic momentum, and that this momentum depends on circumstances external to the conductor, then both induction of currents and electromagnetic attractions may be proved by mechanical reasoning.

What I have called electromagnetic momentum is the same quantity which is called by FARADAY* the electrotonic state of the circuit, every change of which involves the action of an electromotive force, just as change of momentum involves the action of mechanical force.

If, therefore, the phenomena described by FARADAY in the Ninth Series of his Experimental Researches were the only known facts about electric currents, the laws of AMPÈRE relating to the attraction of conductors carrying currents, as well as those of FARADAY about the mutual induction of currents, might be deduced by mechanical reasoning.

In order to bring these results within the range of experimental verification, I shall next investigate the case of a single current, of two currents, and of the six currents in the electric balance, so as to enable the experimenter to determine the values of L, M, N.

* Experimental Researches, Series I. 60, &c.

Case of a single Circuit.

(35) The equation of the current x in a circuit whose resistance is R, and whose coefficient of self-induction is L, acted on by an external electromotive force ξ, is

$$\xi - Rx = \frac{d}{dt} Lx. \quad \ldots \ldots \ldots \quad (13)$$

When ξ is constant, the solution is of the form

$$x = b + (a-b)e^{-\frac{R}{L}t},$$

where a is the value of the current at the commencement, and b is its final value.

The total quantity of electricity which passes in time t, where t is great, is

$$\int_0^t x dt = bt + (a-b)\frac{L}{R}. \quad \ldots \ldots \ldots \quad (14)$$

The value of the integral of x^2 with respect to the time is

$$\int_0^t x^2 dt = b^2 t + (a-b)\frac{L}{R}\left(\frac{3b+a}{2}\right). \quad \ldots \ldots \ldots \quad (15)$$

The actual current changes gradually from the initial value a to the final value b, but the values of the integrals of x and x^2 are the same as if a steady current of intensity $\frac{1}{2}(a+b)$ were to flow for a time $2\frac{L}{R}$, and were then succeeded by the steady current b. The time $2\frac{L}{R}$ is generally so minute a fraction of a second, that the effects on the galvanometer and dynamometer may be calculated as if the impulse were instantaneous.

If the circuit consists of a battery and a coil, then, when the circuit is first completed, the effects are the same as if the current had only half its final strength during the time $2\frac{L}{R}$. This diminution of the current, due to induction, is sometimes called the counter-current.

(36) If an additional resistance r is suddenly thrown into the circuit, as by breaking contact, so as to force the current to pass through a thin wire of resistance r, then the original current is $a = \frac{\xi}{R}$, and the final current is $b = \frac{\xi}{R+r}$.

The current of induction is then $\frac{1}{2}\xi\frac{2R+r}{R(R+r)}$, and continues for a time $2\frac{L}{R+r}$. This current is greater than that which the battery can maintain in the two wires R and r, and may be sufficient to ignite the thin wire r.

When contact is broken by separating the wires in air, this additional resistance is given by the interposed air, and since the electromotive force across the new resistance is very great, a spark will be forced across.

If the electromotive force is of the form $E\sin pt$, as in the case of a coil revolving in a magnetic field, then

$$x=\frac{E}{\varrho}\sin(pt-\alpha),$$

where $\varrho^2=R^2+L^2p^2$, and $\tan\alpha=\frac{Lp}{R}$.

Case of two Circuits.

(37) Let R be the primary circuit and S the secondary circuit, then we have a case similar to that of the induction coil.

The equations of currents are those marked A and B, and we may here assume L, M, N as constant because there is no motion of the conductors. The equations then become

$$\left.\begin{aligned}Rx+L\frac{dx}{dt}+M\frac{dy}{dt}&=\xi,\\ Sy+M\frac{dx}{dt}+N\frac{dy}{dt}&=0.\end{aligned}\right\} \quad \ldots \quad (13^*)$$

To find the total quantity of electricity which passes, we have only to integrate these equations with respect to t; then if x_0, y_0 be the strengths of the currents at time 0, and x_1, y_1 at time t, and if X, Y be the quantities of electricity passed through each circuit during time t,

$$\left.\begin{aligned}X&=\frac{1}{R}\{\xi t+L(x_0-x_1)+M(y_0-y_1)\},\\ Y&=\frac{1}{S}\{M(x_0-x_1)+N(y_0-y_1)\}.\end{aligned}\right\} \quad \ldots \quad (14^*)$$

When the circuit R is completed, then the total currents up to time t, when t is great, are found by making

$$x_0=0,\quad x_1=\frac{\xi}{R},\quad y_0=0,\quad y_1=0;$$

then

$$X=x_1\left(t-\frac{L}{R}\right),\quad Y=-\frac{M}{S}x_1. \quad \ldots \quad (15^*)$$

The value of the total counter-current in R is therefore independent of the secondary circuit, and the induction current in the secondary circuit depends only on M, the coefficient of induction between the coils, S the resistance of the secondary coil, and x_1 the final strength of the current in R.

When the electromotive force ξ ceases to act, there is an extra current in the primary circuit, and a positive induced current in the secondary circuit, whose values are equal and opposite to those produced on making contact.

(38) All questions relating to the total quantity of transient currents, as measured by the impulse given to the magnet of the galvanometer, may be solved in this way without the necessity of a complete solution of the equations. The heating effect of

the current, and the impulse it gives to the suspended coil of WEBER'S dynamometer, depend on the square of the current at every instant during the short time it lasts. Hence we must obtain the solution of the equations, and from the solution we may find the effects both on the galvanometer and dynamometer; and we may then make use of the method of WEBER for estimating the intensity and duration of a current uniform while it lasts which would produce the same effects.

(39) Let n_1, n_2 be the roots of the equation

$$(\mathrm{LN}-\mathrm{M}^2)n^2+(\mathrm{RN}+\mathrm{LS})n+\mathrm{RS}=0, \quad \ldots \quad (16)$$

and let the primary coil be acted on by a constant electromotive force Rc, so that c is the constant current it could maintain; then the complete solution of the equations for making contact is

$$x=\frac{c}{\mathrm{S}}\frac{n_1 n_2}{n_1-n_2}\left\{\left(\frac{\mathrm{S}}{n_1}+\mathrm{N}\right)e^{n_1 t}-\left(\frac{\mathrm{S}}{n_2}+\mathrm{N}\right)e^{n_2 t}+\mathrm{S}\frac{n_1-n_2}{n_1 n_2}\right\}, \quad \ldots \quad (17)$$

$$y=\frac{c\mathrm{M}}{\mathrm{S}}\frac{n_1 n_2}{n_1-n_2}\{e^{n_1 t}-e^{n_2 t}\}. \quad \ldots \quad (18)$$

From these we obtain for calculating the impulse on the dynamometer,

$$\int x^2 dt=c^2\left\{t-\tfrac{3}{2}\frac{\mathrm{L}}{\mathrm{R}}-\tfrac{1}{2}\frac{\mathrm{M}^2}{\mathrm{RN}+\mathrm{LS}}\right\}, \quad \ldots \quad (19)$$

$$\int y^2 dt=c^2\tfrac{1}{2}\frac{\mathrm{M}^2\mathrm{R}}{\mathrm{S}(\mathrm{RN}+\mathrm{LS})}. \quad \ldots \quad (20)$$

The effects of the current in the secondary coil on the galvanometer and dynamometer are the same as those of a uniform current

$$-\tfrac{1}{2}c\frac{\mathrm{MR}}{\mathrm{RN}+\mathrm{LS}}$$

for a time

$$2\left(\frac{\mathrm{L}}{\mathrm{R}}+\frac{\mathrm{N}}{\mathrm{S}}\right).$$

(40) The equation between work and energy may be easily verified. The work done by the electromotive force is

$$\xi\int x dt=c^2(\mathrm{R}t-\mathrm{L}).$$

Work done in overcoming resistance and producing heat,

$$\mathrm{R}\int x^2 dt+\mathrm{S}\int y^2 dt=c^2(\mathrm{R}t-\tfrac{3}{2}\mathrm{L}).$$

Energy remaining in the system,

$$=\tfrac{1}{2}c^2\mathrm{L}.$$

(41) If the circuit R is suddenly and completely interrupted while carrying a current c, then the equation of the current in the secondary coil would be

$$y=c\frac{\mathrm{M}}{\mathrm{N}}e^{-\frac{\mathrm{S}}{\mathrm{N}}t}.$$

This current begins with a value $c\frac{\mathrm{M}}{\mathrm{N}}$, and gradually disappears.

The total quantity of electricity is $c\frac{M}{S}$, and the value of $\int y^2dt$ is $c^2\frac{M^2}{2SN}\cdot$

The effects on the galvanometer and dynamometer are equal to those of a uniform current $\frac{1}{2}c\frac{M}{N}$ for a time $2\frac{N}{S}\cdot$

The heating effect is therefore greater than that of the current on making contact.

(42) If an electromotive force of the form $\xi=E\cos pt$ acts on the circuit R, then if the circuit S is removed, the value of x will be

$$x=\frac{E}{A}\sin(pt-\alpha),$$

where

$$A^2=R^2+L^2p^2,$$

and

$$\tan\alpha=\frac{Lp}{R}\cdot$$

The effect of the presence of the circuit S in the neighbourhood is to alter the value of A and α, to that which they would be if R become

$$R+p^2\frac{MS}{S^2+p^2N^2},$$

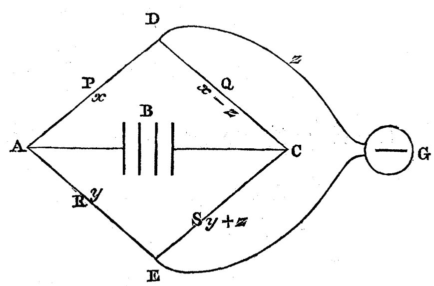
and L became

$$L-p^2\frac{MN}{S^2+p^2N^2}\cdot$$

Hence the effect of the presence of the circuit S is to increase the apparent resistance and diminish the apparent self-induction of the circuit R.

On the Determination of Coefficients of Induction by the Electric Balance.

(43) The electric balance consists of six conductors joining four points, A C D E, two and two. One pair, A C, of these points is connected through the battery B. The opposite pair, D E, is connected through the galvanometer G. Then if the resistances of the four remaining conductors are represented by P, Q, R, S, and the currents in them by x, $x-z$, y, and $y+z$, the current through G will be z. Let the potentials at the four points be A, C, D, E. Then the conditions of steady currents may be found from the equations



$$\left.\begin{aligned}Px&=A-D & Q(x-z)&=D-C,\\ Ry&=A-E & S(y+z)&=E-C,\\ Gz&=D-E & B(x+y)&=-A+C+F.\end{aligned}\right\} \quad \text{(21)}$$

Solving these equations for z, we find

$$z\left\{\frac{1}{P}+\frac{1}{Q}+\frac{1}{R}+\frac{1}{S}+B\left(\frac{1}{P}+\frac{1}{R}\right)\left(\frac{1}{Q}+\frac{1}{S}\right)+G\left(\frac{1}{P}+\frac{1}{Q}\right)\left(\frac{1}{R}+\frac{1}{S}\right)+\frac{BG}{PQRS}(P+Q+R+S)\right\}=F\left(\frac{1}{PS}-\frac{1}{QR}\right)\cdot \quad (22)$$

In this expression F is the electromotive force of the battery, z the current through the galvanometer when it has become steady. P, Q, R, S the resistances in the four arms. B that of the battery and electrodes, and G that of the galvanometer.

(44) If PS=QR, then $z=0$, and there will be no steady current, but a transient current through the galvanometer may be produced on making or breaking circuit on account of induction, and the indications of the galvanometer may be used to determine the coefficients of induction, provided we understand the actions which take place.

We shall suppose PS=QR, so that the current z vanishes when sufficient time is allowed, and

$$x(\mathrm{P}+\mathrm{Q})=y(\mathrm{R}+\mathrm{S})=\frac{\mathrm{F}(\mathrm{P}+\mathrm{Q})(\mathrm{R}+\mathrm{S})}{(\mathrm{P}+\mathrm{Q})(\mathrm{R}+\mathrm{S})+\mathrm{B}(\mathrm{P}+\mathrm{Q})(\mathrm{R}+\mathrm{S})}.$$

Let the induction coefficients between P, Q, R S, be given by the following Table, the coefficient of induction of P on itself being p, between P and Q, h, and so on.

	P	Q	R	S
P	p	h	k	l
Q	h	q	m	n
R	k	m	r	o
S	l	n	o	s

Let g be the coefficient of induction of the galvanometer on itself, and let it be out of the reach of the inductive influence of P, Q, R, S (as it must be in order to avoid direct action of P, Q, R, S on the needle). Let X, Y, Z be the integrals of x, y, z with respect to t. At making contact x, y, z are zero. After a time z disappears, and x and y reach constant values. The equations for each conductor will therefore be

$$\left.\begin{array}{l}\mathrm{PX} \quad +(p+h)x+(k+l)y=\int \mathrm{A}dt-\int \mathrm{D}dt,\\ \mathrm{Q}(\mathrm{X}-\mathrm{Z})+(h+q)x+(m+n)y=\int \mathrm{D}dt-\int \mathrm{C}dt,\\ \mathrm{RY} \quad +(k+m)x+(r+o)y=\int \mathrm{A}dt-\int \mathrm{E}dt,\\ \mathrm{S}(\mathrm{Y}+\mathrm{Z})+(l+n)x+(o+s)y=\int \mathrm{E}dt-\int \mathrm{C}dt,\\ \mathrm{GZ}=\int \mathrm{D}td-\int \mathrm{E}dt.\end{array}\right\} \quad \ldots \ldots \quad (24)$$

Solving these equations for Z, we find

$$\left.\begin{array}{l}\mathrm{Z}\left\{\frac{1}{\mathrm{P}}+\frac{1}{\mathrm{Q}}+\frac{1}{\mathrm{R}}+\frac{1}{\mathrm{S}}+\mathrm{B}\left(\frac{1}{\mathrm{P}}+\frac{1}{\mathrm{R}}\right)\left(\frac{1}{\mathrm{Q}}+\frac{1}{\mathrm{S}}\right)+\mathrm{G}\left(\frac{1}{\mathrm{P}}+\frac{1}{\mathrm{Q}}\right)\left(\frac{1}{\mathrm{R}}+\frac{1}{\mathrm{S}}\right)+\frac{\mathrm{BG}}{\mathrm{PQRS}}(\mathrm{P}+\mathrm{Q}+\mathrm{R}+\mathrm{S})\right\}\\ =-\mathrm{F}\frac{1}{\mathrm{PS}}\left\{\frac{p}{\mathrm{P}}-\frac{q}{\mathrm{Q}}-\frac{r}{\mathrm{R}}+\frac{s}{\mathrm{S}}+h\left(\frac{1}{\mathrm{P}}-\frac{1}{\mathrm{Q}}\right)+k\left(\frac{1}{\mathrm{R}}-\frac{1}{\mathrm{P}}\right)+l\left(\frac{1}{\mathrm{R}}+\frac{1}{\mathrm{Q}}\right)-m\left(\frac{1}{\mathrm{P}}+\frac{1}{\mathrm{S}}\right)\right.\\ \left.+n\left(\frac{1}{\mathrm{Q}}-\frac{1}{\mathrm{S}}\right)+o\left(\frac{1}{\mathrm{S}}-\frac{1}{\mathrm{R}}\right)\right\}.\end{array}\right\} \quad (25)$$

(45) Now let the deflection of the galvanometer by the instantaneous current whose intensity is Z be α.

Let the permanent deflection produced by making the ratio of PS to QR, ϱ instead of unity, be θ.

Also let the time of vibration of the galvanometer needle from rest to rest be T.

Then calling the quantity

$$\frac{p}{P}-\frac{q}{Q}-\frac{r}{R}+\frac{s}{S}+h\left(\frac{1}{P}-\frac{1}{Q}\right)+k\left(\frac{1}{R}-\frac{1}{P}\right)+l\left(\frac{1}{R}+\frac{1}{Q}\right)-m\left(\frac{1}{P}+\frac{1}{S}\right)+n\left(\frac{1}{Q}-\frac{1}{S}\right)+o\left(\frac{1}{S}-\frac{1}{R}\right)=\tau, \quad (26)$$

we find

$$\frac{Z}{z}=\frac{2\sin\frac{1}{2}\alpha}{\tan\theta}\frac{T}{\pi}=\frac{\tau}{1-\varrho}. \quad (27)$$

In determining τ by experiment, it is best to make the alteration of resistance in one of the arms by means of the arrangement described by Mr. JENKIN in the Report of the British Association for 1863, by which any value of ϱ from 1 to 1·01 can be accurately measured.

We observe (α) the greatest deflection due to the impulse of induction when the galvanometer is in circuit, when the connexions are made, and when the resistances are so adjusted as to give no permanent current.

We then observe (β) the greatest deflection produced by the permanent current when the resistance of one of the arms is increased in the ratio of 1 to ϱ, the galvanometer not being in circuit till a little while after the connexion is made with the battery.

In order to eliminate the effects of resistance of the air, it is best to vary ϱ till $\beta=2\alpha$ nearly; then

$$\tau=T\frac{1}{\pi}(1-\varrho)\frac{2\sin\frac{1}{2}\alpha}{\tan\frac{1}{2}\beta}. \quad (28)$$

If all the arms of the balance except P consist of resistance coils of very fine wire of no great length and doubled before being coiled, the induction coefficients belonging to these coils will be insensible, and τ will be reduced to $\frac{p}{P}$. The electric balance therefore affords the means of measuring the self-induction of any circuit whose resistance is known.

(46) It may also be used to determine the coefficient of induction between two circuits, as for instance, that between P and S which we have called m; but it would be more convenient to measure this by directly measuring the current, as in (37), without using the balance. We may also ascertain the equality of $\frac{p}{P}$ and $\frac{q}{Q}$ by there being no current of induction, and thus, when we know the value of p, we may determine that of q by a more perfect method than the comparison of deflections.

Exploration of the Electromagnetic Field.

(47) Let us now suppose the primary circuit A to be of invariable form, and let us explore the electromagnetic field by means of the secondary circuit B, which we shall suppose to be variable in form and position.

We may begin by supposing B to consist of a short straight conductor with its extremities sliding on two parallel conducting rails, which are put in connexion at some distance from the sliding-piece.

Then, if sliding the moveable conductor in a given direction increases the value of M, a negative electromotive force will act in the circuit B, tending to produce a negative current in B during the motion of the sliding-piece.

If a current be kept up in the circuit B, then the sliding-piece will itself tend to move in that direction, which causes M to increase. At every point of the field there will always be a certain direction such that a conductor moved in that direction does not experience any electromotive force in whatever direction its extremities are turned. A conductor carrying a current will experience no mechanical force urging it in that direction or the opposite.

This direction is called the direction of the line of magnetic force through that point.

Motion of a conductor across such a line produces electromotive force in a direction perpendicular to the line and to the direction of motion, and a conductor carrying a current is urged in a direction perpendicular to the line and to the direction of the current.

(48) We may next suppose B to consist of a very small plane circuit capable of being placed in any position and of having its plane turned in any direction. The value of M will be greatest when the plane of the circuit is perpendicular to the line of magnetic force. Hence if a current is maintained in B it will tend to set itself in this position, and will of itself indicate, like a magnet, the direction of the magnetic force.

On Lines of Magnetic Force.

(49) Let any surface be drawn, cutting the lines of magnetic force, and on this surface let any system of lines be drawn at small intervals, so as to lie side by side without cutting each other. Next, let any line be drawn on the surface cutting all these lines, and let a second line be drawn near it, its distance from the first being such that the value of M for each of the small spaces enclosed between these two lines and the lines of the first system is equal to unity.

In this way let more lines be drawn so as to form a second system, so that the value of M for every reticulation formed by the intersection of the two systems of lines is unity.

Finally, from every point of intersection of these reticulations let a line be drawn through the field, always coinciding in direction with the direction of magnetic force.

(50) In this way the whole field will be filled with lines of magnetic force at regular intervals, and the properties of the electromagnetic field will be completely expressed by them.

For, 1st, If any closed curve be drawn in the field, the value of M for that curve will be expressed by the *number* of lines of force which *pass through* that closed curve.

2ndly. If this curve be a conducting circuit and be moved through the field, an electromotive force will act in it, represented by the rate of decrease of the number of lines passing through the curve.

3rdly. If a current be maintained in the circuit, the conductor will be acted on by forces tending to move it so as to increase the number of lines passing through it, and the amount of work done by these forces is equal to the current in the circuit multiplied by the number of additional lines.

4thly. If a small plane circuit be placed in the field, and be free to turn, it will place its plane perpendicular to the lines of force. A small magnet will place itself with its axis in the direction of the lines of force.

5thly. If a long uniformly magnetized bar is placed in the field, each pole will be acted on by a force in the direction of the lines of force. The number of lines of force passing through unit of area is equal to the force acting on a unit pole multiplied by a coefficient depending on the magnetic nature of the medium, and called the coefficient of magnetic induction.

In fluids and isotropic solids the value of this coefficient μ is the same in whatever direction the lines of force pass through the substance, but in crystallized, strained, and organized solids the value of μ may depend on the direction of the lines of force with respect to the axes of crystallization, strain, or growth.

In all bodies μ is affected by temperature, and in iron it appears to diminish as the intensity of the magnetization increases.

On Magnetic Equipotential Surfaces.

(51) If we explore the field with a uniformly magnetized bar, so long that one of its poles is in a very weak part of the magnetic field, then the magnetic forces will perform work on the other pole as it moves about the field.

If we start from a given point, and move this pole from it to any other point, the work performed will be independent of the path of the pole between the two points; provided that no electric current passes between the different paths pursued by the pole.

Hence, when there are no electric currents but only magnets in the field, we may draw a series of surfaces such that the work done in passing from one to another shall be constant whatever be the path pursued between them. Such surfaces are called Equipotential Surfaces, and in ordinary cases are perpendicular to the Lines of magnetic force.

If these surfaces are so drawn that, when a unit pole passes from any one to the next in order, unity of work is done, then the work done in any motion of a magnetic pole will be measured by the strength of the pole multiplied by the number of surfaces which it has passed through in the positive direction.

(52) If there are circuits carrying electric currents in the field, then there will still be equipotential surfaces in the parts of the field external to the conductors carrying the currents, but the work done on a unit pole in passing from one to another will depend on the number of times which the path of the pole circulates round any of these currents. Hence the potential in each surface will have a series of values in arithmetical progression, differing by the work done in passing completely round one of the currents in the field.

The equipotential surfaces will not be continuous closed surfaces, but some of them will be limited sheets, terminating in the electric circuit as their common edge or boundary. The number of these will be equal to the amount of work done on a unit pole in going round the current, and this by the ordinary measurement $=4\pi\gamma$, where γ is the value of the current.

These surfaces, therefore, are connected with the electric current as soap-bubbles are connected with a ring in M. PLATEAU'S experiments. Every current γ has $4\pi\gamma$ surfaces attached to it. These surfaces have the current for their common edge, and meet it at equal angles. The form of the surfaces in other parts depends on the presence of other currents and magnets, as well as on the shape of the circuit to which they belong.

PART III.—GENERAL EQUATIONS OF THE ELECTROMAGNETIC FIELD.

(53.) Let us assume three rectangular directions in space as the axes of x, y, and z, and let all quantities having direction be expressed by their components in these three directions.

Electrical Currents (p, q, r).

(54) An electrical current consists in the transmission of electricity from one part of a body to another. Let the quantity of electricity transmitted in unit of time across unit of area perpendicular to the axis of x be called p, then p is the component of the current at that place in the direction of x.

We shall use the letters p, q, r to denote the components of the current per unit of area in the directions of x, y, z.

Electrical Displacements (f, g, h).

(55) Electrical displacement consists in the opposite electrification of the sides of a molecule or particle of a body which may or may not be accompanied with transmission through the body. Let the quantity of electricity which would appear on the faces $dy.dz$ of an element dx, dy, dz cut from the body be $f.dy.dz$, then f is the component of electric displacement parallel to x. We shall use f, g, h to denote the electric displacements parallel to x, y, z respectively.

The variations of the electrical displacement must be added to the currents p, q, r to get the total motion of electricity, which we may call p', q', r', so that

$$\left.\begin{aligned} p' &= p+\frac{df}{dt}, \\ q' &= q+\frac{dg}{dt}, \\ r' &= r+\frac{dh}{dt}, \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad \text{(A)}$$

Electromotive Force (P, Q, R).

(56) Let P, Q, R represent the components of the electromotive force at any point. Then P represents the difference of potential per unit of length in a conductor

placed in the direction of x at the given point. We may suppose an indefinitely short wire placed parallel to x at a given point and touched, during the action of the force P, by two small conductors, which are then insulated and removed from the influence of the electromotive force. The value of P might then be ascertained by measuring the charge of the conductors.

Thus if l be the length of the wire, the difference of potential at its ends will be Pl, and if C be the capacity of each of the small conductors the charge on each will be $\frac{1}{2}$CPl. Since the capacities of moderately large conductors, measured on the electromagnetic system, are exceedingly small, ordinary electromotive forces arising from electromagnetic actions could hardly be measured in this way. In practice such measurements are always made with long conductors, forming closed or nearly closed circuits.

Electromagnetic Momentum (F, G, H).

(57) Let F, G, H represent the components of electromagnetic momentum at any point of the field, due to any system of magnets or currents.

Then F is the total impulse of the electromotive force in the direction of x that would be generated by the removal of these magnets or currents from the field, that is, if P be the electromotive force at any instant during the removal of the system

$$\mathrm{F}=\int \mathrm{P}\,dt.$$

Hence the part of the electromotive force which depends on the motion of magnets or currents in the field, or their alteration of intensity, is

$$\mathrm{P}=-\frac{d\mathrm{F}}{dt},\quad \mathrm{Q}=-\frac{d\mathrm{G}}{dt},\quad \mathrm{R}=-\frac{d\mathrm{H}}{dt}. \qquad (29)$$

Electromagnetic Momentum of a Circuit.

(58) Let s be the length of the circuit, then if we integrate

$$\int\left(\mathrm{F}\frac{dx}{ds}+\mathrm{G}\frac{dy}{ds}+\mathrm{H}\frac{dz}{ds}\right)ds \qquad (30)$$

round the circuit, we shall get the total electromagnetic momentum of the circuit, or the number of lines of magnetic force which pass through it, the variations of which measure the total electromotive force in the circuit. This electromagnetic momentum is the same thing to which Professor FARADAY has applied the name of the Electrotonic State.

If the circuit be the boundary of the elementary area $dy\,dz$, then its electromagnetic momentum is

$$\left(\frac{d\mathrm{H}}{dy}-\frac{d\mathrm{G}}{dz}\right)dy\,dz,$$

and this is the number of lines of magnetic force which pass through the area $dy\,dz$.

Magnetic Force (α, β, γ).

(59) Let α, β, γ represent the force acting on a unit magnetic pole placed at the given point resolved in the directions of x, y, and z.

Coefficient of Magnetic Induction (μ).

(60) Let μ be the ratio of the magnetic induction in a given medium to that in air under an equal magnetizing force, then the number of lines of force in unit of area perpendicular to x will be $\mu\alpha$ (μ is a quantity depending on the nature of the medium, its temperature, the amount of magnetization already produced, and in crystalline bodies varying with the direction).

(61) Expressing the electric momentum of small circuits perpendicular to the three axes in this notation, we obtain the following

Equations of Magnetic Force.

$$\left.\begin{aligned} \mu\alpha &= \frac{dH}{dy}-\frac{dG}{dz}, \\ \mu\beta &= \frac{dF}{dz}-\frac{dH}{dx}, \\ \mu\gamma &= \frac{dG}{dx}-\frac{dF}{dy}. \end{aligned}\right\} \quad \text{(B)}$$

Equations of Currents.

(62) It is known from experiment that the motion of a magnetic pole in the electromagnetic field in a closed circuit cannot generate work unless the circuit which the pole describes passes round an electric current. Hence, except in the space occupied by the electric currents,

$$\alpha dx+\beta dy+\gamma dz=d\phi \quad (31)$$

a complete differential of ϕ, the magnetic potential.

The quantity ϕ may be susceptible of an indefinite number of distinct values, according to the number of times that the exploring point passes round electric currents in its course, the difference between successive values of ϕ corresponding to a passage completely round a current of strength c being $4\pi c$.

Hence if there is no electric current,

$$\frac{}{dy}-\frac{d\beta}{dz}=0\,;$$

but if there is a current p',

$$\frac{d\gamma}{dy}-\frac{d\beta}{dz}=4\pi p'.$$

Similarly,

$$\left.\begin{aligned} \frac{d\gamma}{dy}-\frac{d\beta}{dz}&=4\pi p'. \\ \frac{d\alpha}{dz}-\frac{d\gamma}{dx}&=4\pi q', \\ \frac{d\beta}{dx}-\frac{d\alpha}{dy}&=4\pi r'. \end{aligned}\right\} \quad \text{(C)}$$

We may call these the Equations of Currents.

Electromotive Force in a Circuit.

(63) Let ξ be the electromotive force acting round the circuit A, then

$$\xi=\int\left(\mathrm{P}\frac{dx}{ds}+\mathrm{Q}\frac{dy}{ds}+\mathrm{R}\frac{dz}{ds}\right)ds, \quad \ldots \ldots \ldots \quad (32)$$

where ds is the element of length, and the integration is performed round the circuit.

Let the forces in the field be those due to the circuits A and B, then the electromagnetic momentum of A is

$$\int\left(\mathrm{F}\frac{dx}{ds}+\mathrm{G}\frac{dy}{ds}+\mathrm{H}\frac{dz}{ds}\right)ds=\mathrm{L}u+\mathrm{M}v, \quad \ldots \ldots \ldots \quad (33)$$

where u and v are the currents in A and B, and

$$\xi=-\frac{d}{dt}(\mathrm{L}u+\mathrm{M}v). \quad \ldots \ldots \ldots \quad (34)$$

Hence, if there is no motion of the circuit A,

$$\left.\begin{aligned}\mathrm{P}&=-\frac{d\mathrm{F}}{dt}-\frac{d\Psi}{dx},\\ \mathrm{Q}&=-\frac{d\mathrm{G}}{dt}-\frac{d\Psi}{dy},\\ \mathrm{R}&=-\frac{d\mathrm{H}}{dt}-\frac{d\Psi}{dz},\end{aligned}\right\} \quad \ldots \ldots \ldots \quad (35)$$

where Ψ is a function of x, y, z, and t, which is indeterminate as far as regards the solution of the above equations, because the terms depending on it will disappear on integrating round the circuit. The quantity Ψ can always, however, be determined in any particular case when we know the actual conditions of the question. The physical interpretation of Ψ is, that it represents the *electric potential* at each point of space.

Electromotive Force on a Moving Conductor.

(64) Let a short straight conductor of length a, parallel to the axis of x, move with a velocity whose components are $\frac{dx}{dt}$, $\frac{dy}{dt}$, $\frac{dz}{dt}$, and let its extremities slide along two parallel conductors with a velocity $\frac{ds}{dt}$. Let us find the alteration of the electromagnetic momentum of the circuit of which this arrangement forms a part.

In unit of time the moving conductor has travelled distances $\frac{dx}{dt}$, $\frac{dy}{dt}$, $\frac{dz}{dt}$ along the directions of the three axes, and at the same time the lengths of the parallel conductors included in the circuit have each been increased by $\frac{ds}{dt}$.

Hence the quantity

$$\int\left(\mathrm{F}\frac{dx}{ds}+\mathrm{G}\frac{dy}{ds}+\mathrm{H}\frac{dz}{ds}\right)ds$$

will be increased by the following increments,

$$a\left(\frac{dF}{dx}\frac{dx}{dt}+\frac{dF}{dy}\frac{dy}{dt}+\frac{dF}{dz}\frac{dz}{dt}\right), \text{ due to motion of conductor,}$$

$$-a\frac{ds}{dt}\left(\frac{dF}{dx}\frac{dx}{ds}+\frac{dG}{dx}\frac{dy}{ds}+\frac{dH}{dx}\frac{dz}{ds}\right), \text{ due to lengthening of circuit.}$$

The total increment will therefore be

$$a\left(\frac{dF}{dy}-\frac{dG}{dx}\right)\frac{dy}{dt}-a\left(\frac{dH}{dx}-\frac{dF}{dz}\right)\frac{dz}{dt};$$

or, by the equations of Magnetic Force (8),

$$-a\left(\mu\gamma\frac{dy}{dt}-\mu\beta\frac{dz}{dt}\right).$$

If P is the electromotive force in the moving conductor parallel to x referred to unit of length, then the actual electromotive force is Pa; and since this is measured by the decrement of the electromagnetic momentum of the circuit, the electromotive force due to motion will be

$$P=\mu\gamma\frac{dy}{dt}-\mu\beta\frac{dz}{dt}. \quad \ldots\ldots\ldots \quad (36)$$

(65) The complete equations of electromotive force on a moving conductor may now be written as follows:—

Equations of Electromotive Force.

$$\left.\begin{aligned} P&=\mu\left(\gamma\frac{dy}{dt}-\beta\frac{dz}{dt}\right)-\frac{dF}{dt}-\frac{d\Psi}{dx}, \\ Q&=\mu\left(\alpha\frac{dz}{dt}-\gamma\frac{dx}{dt}\right)-\frac{dG}{dt}-\frac{d\Psi}{dy}, \\ R&=\mu\left(\beta\frac{dx}{dt}-\alpha\frac{dy}{dt}\right)-\frac{dH}{dt}-\frac{d\Psi}{dz}. \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (D)$$

The first term on the right-hand side of each equation represents the electromotive force arising from the motion of the conductor itself. This electromotive force is perpendicular to the direction of motion and to the lines of magnetic force; and if a parallelogram be drawn whose sides represent in direction and magnitude the velocity of the conductor and the magnetic induction at that point of the field, then the area of the parallelogram will represent the electromotive force due to the motion of the conductor, and the direction of the force is perpendicular to the plane of the parallelogram.

The second term in each equation indicates the effect of changes in the position or strength of magnets or currents in the field.

The third term shows the effect of the electric potential Ψ. It has no effect in causing a circulating current in a closed circuit. It indicates the existence of a force urging the electricity to or from certain definite points in the field.

Electric Elasticity.

(66) When an electromotive force acts on a dielectric, it puts every part of the dielectric into a polarized condition, in which its opposite sides are oppositely electrified. The amount of this electrification depends on the electromotive force and on the nature of the substance, and, in solids having a structure defined by axes, on the direction of the electromotive force with respect to these axes. In isotropic substances, if k is the ratio of the electromotive force to the electric displacement, we may write the

Equations of Electric Elasticity,

$$\left.\begin{aligned} P&=kf, \\ Q&=kg, \\ R&=kh. \end{aligned}\right\} \quad \text{(E)}$$

Electric Resistance.

(67) When an electromotive force acts on a conductor it produces a current of electricity through it. This effect is additional to the electric displacement already considered. In solids of complex structure, the relation between the electromotive force and the current depends on their direction through the solid. In isotropic substances, which alone we shall here consider, if ϱ is the specific resistance referred to unit of volume, we may write the

Equations of Electric Resistance,

$$\left.\begin{aligned} P&=-\varrho p, \\ Q&=-\varrho q, \\ R&=-\varrho r. \end{aligned}\right\} \quad \text{(F)}$$

Electric Quantity.

(68) Let e represent the quantity of free positive electricity contained in unit of volume at any part of the field, then, since this arises from the electrification of the different parts of the field not neutralizing each other, we may write the

Equation of Free Electricity,

$$e+\frac{df}{dx}+\frac{dg}{dy}+\frac{dh}{dz}=0. \quad \text{(G)}$$

(69) If the medium conducts electricity, then we shall have another condition, which may be called, as in hydrodynamics, the

Equation of Continuity,

$$\frac{de}{dt}+\frac{dp}{dx}+\frac{dq}{dy}+\frac{dr}{dz}=0. \quad \text{(H)}$$

(70) In these equations of the electromagnetic field we have assumed twenty variable

quantities, namely,

For Electromagnetic Momentum	F	G	H
„ Magnetic Intensity	α	β	γ
„ Electromotive Force	P	Q	R
„ Current due to true conduction	p	q	r
„ Electric Displacement	f	g	h
„ Total Current (including variation of displacement)	p'	q'	r'
„ Quantity of free Electricity	e		
„ Electric Potential	Ψ		

Between these twenty quantities we have found twenty equations, viz.

Three equations of Magnetic Force	(B)
„ Electric Currents	(C)
„ Electromotive Force	(D)
„ Electric Elasticity	(E)
„ Electric Resistance	(F)
„ Total Currents	(A)
One equation of Free Electricity	(G)
„ Continuity	(H)

These equations are therefore sufficient to determine all the quantities which occur in them, provided we know the conditions of the problem. In many questions, however, only a few of the equations are required.

Intrinsic Energy of the Electromagnetic Field.

(71) We have seen (33) that the intrinsic energy of any system of currents is found by multiplying half the current in each circuit into its electromagnetic momentum. This is equivalent to finding the integral

$$E=\tfrac{1}{2}\Sigma(Fp'+Gq'+Hr')dV \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

over all the space occupied by currents, where p, q, r are the components of currents, and F, G, H the components of electromagnetic momentum.

Substituting the values of p', q', r' from the equations of Currents (C), this becomes

$$\frac{1}{8\pi}\Sigma\left\{F\left(\frac{d\gamma}{dy}-\frac{d\beta}{dz}\right)+G\left(\frac{d\alpha}{dz}-\frac{d\gamma}{dx}\right)+H\left(\frac{d\beta}{dx}-\frac{d\alpha}{dy}\right)\right\}dV.$$

Integrating by parts, and remembering that α, β, γ vanish at an infinite distance, the expression becomes

$$\frac{1}{8\pi}\Sigma\left\{\alpha\left(\frac{dH}{dy}-\frac{dG}{dz}\right)+\beta\left(\frac{dF}{dz}-\frac{dH}{dx}\right)+\gamma\left(\frac{dG}{dx}-\frac{dF}{dy}\right)\right\}dV,$$

where the integration is to be extended over all space. Referring to the equations of Magnetic Force (B), p. 482, this becomes

$$E=\frac{1}{8\pi}\Sigma\{\alpha\,.\,\mu\alpha+\beta\,.\,\mu\beta+\gamma\,.\,\mu\gamma\}dV, \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

where α, β, γ are the components of magnetic intensity or the force on a unit magnetic pole, and $\mu\alpha$, $\mu\beta$, $\mu\gamma$ are the components of the quantity of magnetic induction, or the number of lines of force in unit of area.

In isotropic media the value of μ is the same in all directions, and we may express the result more simply by saying that the intrinsic energy of any part of the magnetic field arising from its magnetization is

$$\frac{\mu}{8\pi}\mathrm{I}^2$$

per unit of volume, where I is the magnetic intensity.

(72) Energy may be stored up in the field in a different way, namely, by the action of electromotive force in producing electric displacement. The work done by a variable electromotive force, P, in producing a variable displacement, f, is got by integrating

$$\int \mathrm{P}\,df$$

from $\mathrm{P}=0$ to the given value of P.

Since $\mathrm{P}=kf$, equation (E), this quantity becomes

$$\int kf\,df=\tfrac{1}{2}kf^2=\tfrac{1}{2}\mathrm{P}f.$$

Hence the intrinsic energy of any part of the field, as existing in the form of electric displacement, is

$$\tfrac{1}{2}\Sigma(\mathrm{P}f+\mathrm{Q}g+\mathrm{R}h)d\mathrm{V}.$$

The total energy existing in the field is therefore

$$\mathrm{E}=\Sigma\left\{\frac{1}{8\pi}(\alpha\mu\alpha+\beta\mu\beta+\gamma\mu\gamma)+\tfrac{1}{2}(\mathrm{P}f+\mathrm{Q}g+\mathrm{R}h)\right\}d\mathrm{V}. \quad . \quad . \quad . \quad . \quad (\mathrm{I})$$

The first term of this expression depends on the magnetization of the field, and is explained on our theory by actual motion of some kind. The second term depends on the electric polarization of the field, and is explained on our theory by strain of some kind in an elastic medium.

(73) I have on a former occasion* attempted to describe a particular kind of motion and a particular kind of strain, so arranged as to account for the phenomena. In the present paper I avoid any hypothesis of this kind; and in using such words as electric momentum and electric elasticity in reference to the known phenomena of the induction of currents and the polarization of dielectrics, I wish merely to direct the mind of the reader to mechanical phenomena which will assist him in understanding the electrical ones. All such phrases in the present paper are to be considered as illustrative, not as explanatory.

(74) In speaking of the Energy of the field, however, I wish to be understood literally. All energy is the same as mechanical energy, whether it exists in the form of motion or in that of elasticity, or in any other form. The energy in electromagnetic phenomena is mechanical energy. The only question is, Where does it reside? On the old theories

* "On Physical Lines of Force," Philosophical Magazine, 1861–62.

it resides in the electrified bodies, conducting circuits, and magnets, in the form of an unknown quality called potential energy, or the power of producing certain effects at a distance. On our theory it resides in the electromagnetic field, in the space surrounding the electrified and magnetic bodies, as well as in those bodies themselves, and is in two different forms, which may be described without hypothesis as magnetic polarization and electric polarization, or, according to a very probable hypothesis, as the motion and the strain of one and the same medium.

(75) The conclusions arrived at in the present paper are independent of this hypothesis, being deduced from experimental facts of three kinds:—

1. The induction of electric currents by the increase or diminution of neighbouring currents according to the changes in the lines of force passing through the circuit.

2. The distribution of magnetic intensity according to the variations of a magnetic potential.

3. The induction (or influence) of statical electricity through dielectrics.

We may now proceed to demonstrate from these principles the existence and laws of the mechanical forces which act upon electric currents, magnets, and electrified bodies placed in the electromagnetic field.

PART IV.—MECHANICAL ACTIONS IN THE FIELD.

Mechanical Force on a Moveable Conductor.

(76) We have shown (§§ 34 & 35) that the work done by the electromagnetic forces in aiding the motion of a conductor is equal to the product of the current in the conductor multiplied by the increment of the electromagnetic momentum due to the motion.

Let a short straight conductor of length a move parallel to itself in the direction of x, with its extremities on two parallel conductors. Then the increment of the electromagnetic momentum due to the motion of a will be

$$a\left(\frac{d\mathrm{F}}{dx}\frac{dx}{ds}+\frac{d\mathrm{G}}{dx}\frac{dy}{ds}+\frac{d\mathrm{H}}{dx}\frac{dz}{ds}\right)\delta x.$$

That due to the lengthening of the circuit by increasing the length of the parallel conductors will be

$$-a\left(\frac{d\mathrm{F}}{dx}\frac{dx}{ds}+\frac{d\mathrm{F}}{dy}\frac{dy}{ds}+\frac{d\mathrm{F}}{dz}\frac{dz}{ds}\right)\delta x.$$

The total increment is

$$a\delta x\left\{\frac{dy}{ds}\left(\frac{d\mathrm{G}}{dx}-\frac{d\mathrm{F}}{dy}\right)-\frac{dz}{ds}\left(\frac{d\mathrm{F}}{dz}-\frac{d\mathrm{H}}{dx}\right)\right\},$$

which is by the equations of Magnetic Force (B), p. 482,

$$a\delta x\left(\frac{dy}{ds}\mu\gamma-\frac{dz}{ds}\mu\beta\right).$$

Let X be the force acting along the direction of x per unit of length of the conductor, then the work done is $\mathrm{X}a\delta x$.

Let C be the current in the conductor, and let p', q', r' be its components, then

$$X a\delta = C a\delta x x\left(\frac{dy}{ds}\mu\gamma - \frac{dz}{ds}\mu\beta\right),$$

or

$$\left.\begin{aligned} X &= \mu\gamma q' - \mu\beta r'. \\ Y &= \mu\alpha r' - \mu\gamma p', \\ Z &= \mu\beta p' - \mu\alpha q'. \end{aligned}\right\} \quad \text{(J)}$$

Similarly,

These are the equations which determine the mechanical force acting on a conductor carrying a current. The force is perpendicular to the current and to the lines of force, and is measured by the area of the parallelogram formed by lines parallel to the current and lines of force, and proportional to their intensities.

Mechanical Force on a Magnet.

(77) In any part of the field not traversed by electric currents the distribution of magnetic intensity may be represented by the differential coefficients of a function which may be called the magnetic potential. When there are no currents in the field, this quantity has a single value for each point. When there are currents, the potential has a series of values at each point, but its differential coefficients have only one value, namely,

$$\frac{d\phi}{dx} = \alpha, \quad \frac{d\phi}{dy} = \beta, \quad \frac{d\phi}{dz} = \gamma.$$

Substituting these values of α, β, γ in the expression (equation 38) for the intrinsic energy of the field, and integrating by parts, it becomes

$$-\Sigma\left\{\phi\,\frac{1}{8\pi}\left(\frac{d\mu\alpha}{dx} + \frac{d\mu\beta}{dy} + \frac{d\mu\gamma}{dz}\right)\right\}dV.$$

The expression

$$\Sigma\left(\frac{d\mu\alpha}{dx} + \frac{d\mu\beta}{dy} + \frac{d\mu\gamma}{dz}\right)dV = \Sigma m dV \quad \text{. . . . . . . (39)}$$

indicates the number of lines of magnetic force which have their origin within the space V. Now a magnetic pole is known to us only as the origin or termination of lines of magnetic force, and a unit pole is one which has 4π lines belonging to it, since it produces unit of magnetic intensity at unit of distance over a sphere whose surface is 4π.

Hence if m is the amount of free positive magnetism in unit of volume, the above expression may be written $4\pi m$, and the expression for the energy of the field becomes

$$E = -\Sigma(\tfrac{1}{2}\phi m)dV. \quad \text{. . . . . (40)}$$

If there are two magnetic poles m_1 and m_2 producing potentials ϕ_1 and ϕ_2 in the field, then if m_2 is moved a distance dx, and is urged in that direction by a force X, then the work done is Xdx, and the decrease of energy in the field is

$$d\left(\tfrac{1}{2}(\phi_1+\phi_2)(m_1+m_2)\right),$$

and these must be equal by the principle of Conservation of Energy.

Since the distribution ϕ_1 is determined by m_1, and ϕ_2 by m_2, the quantities $\phi_1 m_1$ and $\phi_2 m_2$ will remain constant.

It can be shown also, as GREEN has proved (Essay, p. 10), that

$$m_1\phi_2 = m_2\phi_1,$$

so that we get

$$Xdx = d(m_2\phi_1),$$

or

$$\left.\begin{aligned} X &= m_2\frac{d\phi_1}{dx} = m_2\alpha_1, \\ &\text{where } \alpha_1 \text{ represents the magnetic intensity due to } m_1. \\ &\text{Similarly, } Y = m_2\beta_1, \\ Z &= m_2\gamma_1. \end{aligned}\right\} \quad \text{(K)}$$

So that a magnetic pole is urged in the direction of the lines of magnetic force with a force equal to the product of the strength of the pole and the magnetic intensity.

(78) If a single magnetic pole, that is one pole of a very long magnet, be placed in the field, the only solution of ϕ is

$$\phi_1 = -\frac{m_1}{\mu}\frac{1}{r}, \qquad (41)$$

where m_1 is the strength of the pole and r the distance from it.

The repulsion between two poles of strength m_1 and m_2 is

$$m_2\frac{d\phi_1}{dr} = \frac{m_1 m_2}{\mu r^2}. \qquad (42)$$

In air or any medium in which $\mu=1$ this is simply $\frac{m_1 m_2}{r^2}$, but in other media the force acting between two given magnetic poles is inversely proportional to the coefficient of magnetic induction for the medium. This may be explained by the magnetization of the medium induced by the action of the poles.

Mechanical Force on an Electrified Body.

(79) If there is no motion or change of strength of currents or magnets in the field, the electromotive force is entirely due to variation of electric potential, and we shall have (§ 65)

$$P = -\frac{d\Psi}{dx}, \quad Q = -\frac{d\Psi}{dy}, \quad R = -\frac{d\Psi}{dz}.$$

Integrating by parts the expression (I) for the energy due to electric displacement, and remembering that P, Q, R vanish at an infinite distance, it becomes

$$\tfrac{1}{2}\Sigma\left\{\Psi\left(\frac{df}{dx}+\frac{dg}{dy}+\frac{dh}{dz}\right)\right\}dV,$$

or by the equation of Free Electricity (G), p. 485,

$$-\tfrac{1}{2}\Sigma(\Psi e)dV.$$

By the same demonstration as was used in the case of the mechanical action on a magnet, it may be shown that the mechanical force on a small body containing a quantity e_2 of free electricity placed in a field whose potential arising from other electrified bodies is Ψ_1, has for components

$$\left.\begin{aligned} X&=e_2\frac{d\Psi_1}{dx}=-P_1e_2,\\ Y&=e_2\frac{d\Psi_1}{dy}=-Q_1e_2,\\ Z&=e_3\frac{d\Psi_1}{dz}=-R_1e_2. \end{aligned}\right\} \quad \text{(D)}$$

So that an electrified body is urged in the direction of the electromotive force with a force equal to the product of the quantity of free electricity and the electromotive force.

If the electrification of the field arises from the presence of a small electrified body containing e_1 of free electrity, the only solution of Ψ_1 is

$$\Psi_1=\frac{k}{4\pi}\frac{e_1}{r}, \quad (43)$$

where r is the distance from the electrified body.

The repulsion between two electrified bodies e_1, e_2 is therefore

$$e_2\frac{d\Psi_1}{dr}=\frac{k}{4\pi}\frac{e_1e_2}{r^2}. \quad (44)$$

Measurement of Electrical Phenomena by Electrostatic Effects.

(80) The quantities with which we have had to do have been hitherto expressed in terms of the Electromagnetic System of measurement, which is founded on the mechanical action between currents. The electrostatic system of measurement is founded on the mechanical action between electrified bodies, and is independent of, and incompatible with, the electromagnetic system; so that the units of the different kinds of quantity have different values according to the system we adopt, and to pass from the one system to the other, a reduction of all the quantities is required.

According to the electrostatic system, the repulsion between two small bodies charged with quantities η_1, η_2 of electricity is

$$\frac{\eta_1\eta_2}{r^2},$$

where r is the distance between them.

Let the relation of the two systems be such that one electromagnetic unit of electricity contains v electrostatic units; then $\eta_1=ve_1$ and $\eta_2=ve_2$, and this repulsion becomes

$$v^2\frac{e_1e_2}{r^2}=\frac{k}{4\pi}\frac{e_1e_2}{r^2} \text{ by equation (44)}, \quad (45)$$

whence k, the coefficient of "electric elasticity" in the medium in which the experiments are made, *i. e.* common air, is related to v, the number of electrostatic units in one electromagnetic unit, by the equation

$$k=4\pi v^2. \quad (46)$$

The quantity v may be determined by experiment in several ways. According to the experiments of MM. WEBER and KOHLRAUSCH,

$$v=310,740,000 \text{ metres per second.}$$

(81) It appears from this investigation, that if we assume that the medium which constitutes the electromagnetic field is, when dielectric, capable of receiving in every part of it an electric polarization, in which the opposite sides of every element into which we may conceive the medium divided are oppositely electrified, and if we also assume that this polarization or electric displacement is proportional to the electromotive force which produces or maintains it, then we can show that electrified bodies in a dielectric medium will act on one another with forces obeying the same laws as are established by experiment.

The energy, by the expenditure of which electrical attractions and repulsions are produced, we suppose to be stored up in the dielectric medium which surrounds the electrified bodies, and not on the surface of those bodies themselves, which on our theory are merely the bounding surfaces of the air or other dielectric in which the true springs of action are to be sought.

Note on the Attraction of Gravitation.

(82) After tracing to the action of the surrounding medium both the magnetic and the electric attractions and repulsions, and finding them to depend on the inverse square of the distance, we are naturally led to inquire whether the attraction of gravitation, which follows the same law of the distance, is not also traceable to the action of a surrounding medium.

Gravitation differs from magnetism and electricity in this; that the bodies concerned are all of the same kind, instead of being of opposite signs, like magnetic poles and electrified bodies, and that the force between these bodies is an attraction and not a repulsion, as is the case between like electric and magnetic bodies.

The lines of gravitating force near two dense bodies are exactly of the same form as the lines of magnetic force near two poles of the same name; but whereas the poles are repelled, the bodies are attracted. Let E be the intrinsic energy of the field surrounding two gravitating bodies M_1, M_2, and let E′ be the intrinsic energy of the field surrounding two magnetic poles m_1, m_2, equal in numerical value to M_1, M_2, and let X be the gravitating force acting during the displacement δx, and X′ the magnetic force,

$$X\delta x=\delta E, \qquad X'\delta x=\delta E';$$

now X and X′ are equal in numerical value, but of opposite signs; so that

$$\delta E=-\delta E',$$

or

$$E=C-E'$$

$$=C-\Sigma\frac{1}{8\pi}(\alpha^2+\beta^2+\gamma^2)dV,$$

where α, β, γ are the components of magnetic intensity. If R be the resultant gravitating force, and R′ the resultant magnetic force at a corresponding part of the field,

$$R=-R', \text{ and } \alpha^2+\beta^2+\gamma^2=R^2=R'^2.$$

Hence

$$E=C-\Sigma\frac{1}{8\pi}R^2dV. \qquad (47)$$

The intrinsic energy of the field of gravitation must therefore be less wherever there is a resultant gravitating force.

As energy is essentially positive, it is impossible for any part of space to have negative intrinsic energy. Hence those parts of space in which there is no resultant force, such as the points of equilibrium in the space between the different bodies of a system, and within the substance of each body, must have an intrinsic energy per unit of volume greater than

$$\frac{1}{8\pi}R^2,$$

where R is the greatest possible value of the intensity of gravitating force in any part of the universe.

The assumption, therefore, that gravitation arises from the action of the surrounding medium in the way pointed out, leads to the conclusion that every part of this medium possesses, when undisturbed, an enormous intrinsic energy, and that the presence of dense bodies influences the medium so as to diminish this energy wherever there is a resultant attraction.

As I am unable to understand in what way a medium can possess such properties, I cannot go any further in this direction in searching for the cause of gravitation.

PART V.—THEORY OF CONDENSERS.

Capacity of a Condenser.

(83) The simplest form of condenser consists of a uniform layer of insulating matter bounded by two conducting surfaces, and its capacity is measured by the quantity of electricity on either surface when the difference of potentials is unity.

Let S be the area of either surface, a the thickness of the dielectric, and k its coefficient of electric elasticity; then on one side of the condenser the potential is Ψ_1, and on the other side Ψ_1+1, and within its substance

$$\frac{d\Psi}{dx}=\frac{1}{a}=kf. \qquad (48)$$

Since $\frac{d\Psi}{dx}$ and therefore f is zero outside the condenser, the quantity of electricity on its first surface $=-Sf$, and on the second $+Sf$. The capacity of the condenser is therefore $Sf=\frac{S}{ak}$ in electromagnetic measure.

Specific Capacity of Electric Induction (D).

(84) If the dielectric of the condenser be air, then its capacity in electrostatic measure is $\frac{S}{4\pi a}$ (neglecting corrections arising from the conditions to be fulfilled at the edges). If the dielectric have a capacity whose ratio to that of air is D, then the capacity of the condenser will be $\frac{DS}{4\pi a}$.

Hence $$D=\frac{k_0}{k}, \quad \ldots \ldots \ldots \ldots \quad (49)$$

where k_0 is the value of k in air, which is taken for unity.

Electric Absorption.

(85) When the dielectric of which the condenser is formed is not a perfect insulator, the phenomena of conduction are combined with those of electric displacement. The condenser, when left charged, gradually loses its charge, and in some cases, after being discharged completely, it gradually acquires a new charge of the same sign as the original charge, and this finally disappears. These phenomena have been described by Professor FARADAY (Experimental Researches, Series XI.) and by Mr. F. JENKIN (Report of Committee of Board of Trade on Submarine Cables), and may be classed under the name of "Electric Absorption."

(86) We shall take the case of a condenser composed of any number of parallel layers of different materials. If a constant difference of potentials between its extreme surfaces is kept up for a sufficient time till a condition of permanent steady flow of electricity is established, then each bounding surface will have a charge of electricity depending on the nature of the substances on each side of it. If the extreme surfaces be now discharged, these internal charges will gradually be dissipated, and a certain charge may reappear on the extreme surfaces if they are insulated, or, if they are connected by a conductor, a certain quantity of electricity may be urged through the conductor during the reestablishment of equilibrium.

Let the thickness of the several layers of the condenser be a_1, a_2, &c.

Let the values of k for these layers be respectively k_1, k_2, k_3, and let

$$a_1k_2+a_2k_2+\&c.=ak, \quad \ldots \ldots \ldots \quad (50)$$

where k is the "electric elasticity" of air, and a is the thickness of an equivalent condenser of air.

Let the resistances of the layers be respectively r_1, r_2, &c., and let r_1+r_2+ &c. $=r$ be the resistance of the whole condenser, to a steady current through it per unit of surface.

Let the electric displacement in each layer be f_1, f_2, &c.

Let the electric current in each layer be p_1, p_2, &c.

Let the potential on the first surface be Ψ_1, and the electricity per unit of surface e_1.

Let the corresponding quantities at the boundary of the first and second surface be Ψ_2 and e_2, and so on. Then by equations (G) and (H),

$$\left.\begin{array}{ll} e_1=-f_1, & \frac{de_1}{dt}=-p_1, \\ e_2=f_1-f_2, & \frac{de_2}{dt}=p_1-p_2, \\ \&c. & \&c. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (51)$$

But by equations (E) and (F),

$$\left.\begin{array}{l} \Psi_1-\Psi_2=a_1k_1f_1=-r_1p_1, \\ \Psi_2-\Psi_3=a_2k_2f_2=-r_2p_2, \\ \quad \&c. \qquad \&c. \qquad \&c. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (52)$$

After the electromotive force has been kept up for a sufficient time the current becomes the same in each layer, and

$$p_1=p_2=\&c.=p=\frac{\Psi}{r},$$

where Ψ is the total difference of potentials between the extreme layers. We have then

and
$$\left.\begin{array}{ll} f_1=-\frac{\Psi}{r}\frac{r_1}{a_1k_1}, & f_2=-\frac{\Psi}{r}\frac{r_2}{a_2k_2}, \ \&c. \\ e_1=\frac{\Psi}{r}\frac{r_1}{a_1k_1}, & e_2=\frac{\Psi}{r}\left(\frac{r_2}{a_2k_2}-\frac{r_1}{a_1k_1}\right), \ \&c. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (53)$$

These are the quantities of electricity on the different surfaces.

(87) Now let the condenser be discharged by connecting the extreme surfaces through a perfect conductor so that their potentials are instantly rendered equal, then the electricity on the extreme surfaces will be altered, but that on the internal surfaces will not have time to escape. The total difference of potentials is now

$$\Psi'=a_1k_1e'_1+a_2k_2(e'_1+e_2)+a_3k_3(e'_1+e_2+e_3), \ \&c.=0, \quad . \quad . \quad . \quad . \quad (54)$$

whence if e'_1 is what e_1 becomes at the instant of discharge,

$$e'_1=\frac{\Psi}{r}\frac{r_1}{a_1k_1}-\frac{\Psi}{ak}=e_1-\frac{\Psi}{ak}. \quad . \quad . \quad . \quad . \quad (55)$$

The instantaneous discharge is therefore $\frac{\Psi}{ak}$, or the quantity which would be discharged by a condenser of air of the equivalent thickness a, and it is unaffected by the want of perfect insulation.

(88) Now let us suppose the connexion between the extreme surfaces broken, and the condenser left to itself, and let us consider the gradual dissipation of the internal charges. Let Ψ' be the difference of potential of the extreme surfaces at any time t; then

$$\Psi'=a_1k_1f_1+a_2k_2f_2+\&c.; \quad . \quad . \quad . \quad . \quad (56)$$

but
$$a_1k_1f_1=-r_1\frac{df_1}{dt},$$

$$a_2k_2f_2=-r_2\frac{df_2}{dt}.$$

Hence $f_1=A_1e^{-\frac{a_1k_1}{r_1}t}$, $f_2=A_2e^{-\frac{a_2k_2}{r_2}t}$, &c.; and by referring to the values of e'_1, e_2, &c., we find

$$\left.\begin{aligned} A_1&=\frac{\Psi}{r}\frac{r_1}{a_1k_1}-\frac{\Psi}{ak}, \\ A_2&=\frac{\Psi}{r}\frac{r_2}{a_2k_2}-\frac{\Psi}{ak}, \\ &\text{\&c.}; \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (57)$$

so that we find for the difference of extreme potentials at any time,

$$\Psi'=\Psi\left\{\left(\frac{r_1}{r}-\frac{a_1k_1}{ak}\right)e^{-\frac{a_1k_1}{r_1}t}+\left(\frac{r_2}{r}-\frac{a_2k_2}{ak}\right)e^{-\frac{a_2k_2}{r_2}t}+\text{\&c.}\right\}. \quad . \quad . \quad . \quad (58)$$

(89) It appears from this result that if all the layers are made of the same substance, Ψ' will be zero always. If they are of different substances, the order in which they are placed is indifferent, and the effect will be the same whether each substance consists of one layer, or is divided into any number of thin layers and arranged in any order among thin layers of the other substances. Any substance, therefore, the parts of which are not mathematically homogeneous, though they may be apparently so, may exhibit phenomena of absorption. Also, since the order of magnitude of the coefficients is the same as that of the indices, the value of Ψ' can never change sign, but must start from zero, become positive, and finally disappear.

(90) Let us next consider the total amount of electricity which would pass from the first surface to the second, if the condenser, after being thoroughly saturated by the current and then discharged, has its extreme surfaces connected by a conductor of resistance R. Let p be the current in this conductor; then, during the discharge,

$$\Psi'=p_1r_1+p_2r_2+\text{\&c.}=pR. \quad . \quad . \quad . \quad . \quad (59)$$

Integrating with respect to the time, and calling q_1, q_2, q the quantities of electricity which traverse the different conductors,

$$q_1r_1+q_2r_2+\text{\&c.}=qR. \quad . \quad . \quad . \quad . \quad (60)$$

The quantities of electricity on the several surfaces will be

$$\begin{aligned} &e'_1-q-q_1, \\ &e_2+q_1-q_2, \\ &\text{\&c.}; \end{aligned}$$

and since at last all these quantities vanish, we find

$$\begin{aligned} q_1&=e'_1-q, \\ q_2&=e'_1+e_2-q; \end{aligned}$$

whence
$$qR=\frac{\Psi}{r}\left(\frac{r_1^2}{a_1k_1}+\frac{r_2^2}{a_2k_2}+\text{\&c.}\right)-\frac{\Psi r}{ak},$$

or
$$q=\frac{\Psi}{akrR}\left\{a_1k_1a_2k_2\left(\frac{r_1}{a_1k_1}-\frac{r_2}{a_2k_2}\right)^2+a_2k_2a_3k_3\left(\frac{r_2}{a_2k_2}-\frac{r_3}{a_3k_3}\right)^2+\text{\&c.}\right\}, \quad . \quad . \quad . \quad (61)$$

a quantity essentially positive; so that, when the primary electrification is in one direction, the secondary discharge is always in the same direction as the primary discharge*.

PART VI.—ELECTROMAGNETIC THEORY OF LIGHT.

(91) At the commencement of this paper we made use of the optical hypothesis of an elastic medium through which the vibrations of light are propagated, in order to show that we have warrantable grounds for seeking, in the same medium, the cause of other phenomena as well as those of light. We then examined electromagnetic phenomena, seeking for their explanation in the properties of the field which surrounds the electrified or magnetic bodies. In this way we arrived at certain equations expressing certain properties of the electromagnetic field. We now proceed to investigate whether these properties of that which constitutes the electromagnetic field, deduced from electromagnetic phenomena alone, are sufficient to explain the propagation of light through the same substance.

(92) Let us suppose that a plane wave whose direction cosines are l, m, n is propagated through the field with a velocity V. Then all the electromagnetic functions will be functions of

$$w=lx+my+nz-\mathrm{V}t.$$

The equations of Magnetic Force (B), p. 482, will become

$$\mu\alpha=m\frac{d\mathrm{H}}{dw}-n\frac{d\mathrm{G}}{dw},$$
$$\mu\beta=n\frac{d\mathrm{F}}{dw}-l\frac{d\mathrm{H}}{dw},$$
$$\mu\gamma=l\frac{d\mathrm{G}}{dw}-m\frac{d\mathrm{F}}{dw}.$$

If we multiply these equations respectively by l, m, n, and add, we find

$$l\mu\alpha+m\mu\beta+n\mu\gamma=0, \quad \ldots \ldots \quad (62)$$

which shows that the direction of the magnetization must be in the plane of the wave.

(93) If we combine the equations of Magnetic Force (B) with those of Electric Currents (C), and put for brevity

$$\frac{d\mathrm{F}}{dx}+\frac{d\mathrm{G}}{dy}+\frac{d\mathrm{H}}{dz}=\mathrm{J}, \text{ and } \frac{d^2}{dx^2}+\frac{d^2}{dy^2}+\frac{d^2}{dz^2}=\nabla^2, \quad \ldots \ldots \quad (63)$$

$$\left.\begin{aligned} 4\pi\mu p'&=\frac{d\mathrm{J}}{dx}-\nabla^2\mathrm{F}, \\ 4\pi\mu q'&=\frac{d\mathrm{J}}{dy}-\nabla^2\mathrm{G}, \\ 4\pi\mu r'&=\frac{d\mathrm{J}}{dz}-\nabla^2\mathrm{H}. \end{aligned}\right\} \quad \ldots \ldots \quad (64)$$

* Since this paper was communicated to the Royal Society, I have seen a paper by M. Gaugain in the Annales de Chimie for 1864, in which he has deduced the phenomena of electric absorption and secondary discharge from the theory of compound condensers.

If the medium in the field is a perfect dielectric there is no true conduction, and the currents p', q', r' are only variations in the electric displacement, or, by the equations of Total Currents (A),

$$p'=\frac{df}{dt}, \qquad q'=\frac{dg}{dt}, \qquad r'=\frac{dh}{dt}. \qquad (65)$$

But these electric displacements are caused by electromotive forces, and by the equations of Electric Elasticity (E),

$$P=kf, \qquad Q=kg, \qquad R=kh. \qquad (66)$$

These electromotive forces are due to the variations either of the electromagnetic or the electrostatic functions, as there is no motion of conductors in the field; so that the equations of electromotive force (D) are

$$\left.\begin{aligned} P&=-\frac{dF}{dt}-\frac{d\Psi}{dx}, \\ Q&=-\frac{dG}{dt}-\frac{d\Psi}{dy}, \\ R&=-\frac{dH}{dt}-\frac{d\Psi}{dz}. \end{aligned}\right\} \qquad (67)$$

(94) Combining these equations, we obtain the following:—

$$\left.\begin{aligned} k\left(\frac{dJ}{dx}-\nabla^2 F\right)+4\pi\mu\left(\frac{d^2F}{dt^2}+\frac{d^2\Psi}{dxdt}\right)&=0, \\ k\left(\frac{dJ}{dy}-\nabla^2 G\right)+4\pi\mu\left(\frac{d^2G}{dt^2}+\frac{d^2\Psi}{dydt}\right)&=0, \\ k\left(\frac{dJ}{dz}-\nabla^2 H\right)+4\pi\mu\left(\frac{d^2H}{dt^2}+\frac{d^2\Psi}{dzdt}\right)&=0. \end{aligned}\right\} \qquad (68)$$

If we differentiate the third of these equations with respect to y, and the second with respect to z, and subtract, J and Ψ disappear, and by remembering the equations (B) of magnetic force, the results may be written

$$\left.\begin{aligned} k\nabla^2\mu\alpha&=4\pi\mu\frac{d^2}{dt^2}\mu\alpha, \\ k\nabla^2\mu\beta&=4\pi\mu\frac{d^2}{dt^2}\mu\beta, \\ k\nabla^2\mu\gamma&=4\pi\mu\frac{d^2}{d^2t}\mu\gamma. \end{aligned}\right\} \qquad (69)$$

(95) If we assume that α, β, γ are functions of $lx+my+nz-Vt=w$, the first equation becomes

$$k\mu\frac{d^2\alpha}{dw^2}=4\pi\mu^2V^2\frac{d^2\alpha}{dw^2}, \qquad (70)$$

or

$$V=\pm\sqrt{\frac{k}{4\pi\mu}}. \qquad (71)$$

The other equations give the same value for V, so that the wave is propagated in either direction with a velocity V.

This wave consists entirely of magnetic disturbances, the direction of magnetization being in the plane of the wave. No magnetic disturbance whose direction of magnetization is not in the plane of the wave can be propagated as a plane wave at all.

Hence magnetic disturbances propagated through the electromagnetic field agree with light in this, that the disturbance at any point is transverse to the direction of propagation, and such waves may have all the properties of polarized light.

(96) The only medium in which experiments have been made to determine the value of k is air, in which $\mu=1$, and therefore, by equation (46),

$$V=v. \qquad (72)$$

By the electromagnetic experiments of MM. WEBER and KOHLRAUSCH*,

$$v=310,740,000 \text{ metres per second}$$

is the number of electrostatic units in one electromagnetic unit of electricity, and this, according to our result, should be equal to the velocity of light in air or vacuum.

The velocity of light in air, by M. FIZEAU'S† experiments, is

$$V=314,858,000;$$

according to the more accurate experiments of M. FOUCAULT‡,

$$V=298,000,000.$$

The velocity of light in the space surrounding the earth, deduced from the coefficient of aberration and the received value of the radius of the earth's orbit, is

$$V=308,000,000.$$

(97) Hence the velocity of light deduced from experiment agrees sufficiently well with the value of v deduced from the only set of experiments we as yet possess. The value of v was determined by measuring the electromotive force with which a condenser of known capacity was charged, and then discharging the condenser through a galvanometer, so as to measure the quantity of electricity in it in electromagnetic measure. The only use made of light in the experiment was to see the instruments. The value of V found by M. FOUCAULT was obtained by determining the angle through which a revolving mirror turned, while the light reflected from it went and returned along a measured course. No use whatever was made of electricity or magnetism.

The agreement of the results seems to show that light and magnetism are affections of the same substance, and that light is an electromagnetic disturbance propagated through the field according to electromagnetic laws.

(98) Let us now go back upon the equations in (94), in which the quantities J and Ψ occur, to see whether any other kind of disturbance can be propagated through the medium depending on these quantities which disappeared from the final equations.

* Leipzig Transactions, vol. v. (1857), p. 260, or POGGENDORFF'S 'Annalen,' Aug. 1856, p. 10.

† Comptes Rendus, vol. xxix. (1849), p. 90. ‡ Ibid. vol. lv. (1862), pp. 501, 792.

If we determine χ from the equation

$$\nabla^2\chi=\frac{d^2\chi}{dx^2}+\frac{d^2\chi}{dy^2}+\frac{d^2\chi}{dz^2}=\mathrm{J}, \quad \ldots \ldots \quad (73)$$

and F′, G′, H′ from the equations

$$\mathrm{F}'=\mathrm{F}-\frac{d\chi}{dx}, \quad \mathrm{G}'=\mathrm{G}-\frac{d\chi}{dy}, \quad \mathrm{H}'=\mathrm{H}-\frac{d\chi}{dz}, \quad \ldots \quad (74)$$

then

$$\frac{d\mathrm{F}'}{dx}+\frac{d\mathrm{G}'}{dy}+\frac{d\mathrm{H}'}{dz}=0, \quad \ldots \ldots \quad (75)$$

and the equations in (94) become of the form

$$k\nabla^2\mathrm{F}'=4\pi\mu\left(\frac{a^2\mathrm{F}'}{dt^2}+\frac{d}{dxdt}\left(\Psi+\frac{d\chi}{dt}\right)\right). \quad \ldots \ldots \quad (76)$$

Differentiating the three equations with respect to x, y, and z, and adding, we find that

$$\Psi=-\frac{d\chi}{dt}+\phi(x, y, z), \quad \ldots \ldots \quad (77)$$

and that

$$\left.\begin{aligned} k\nabla^2\mathrm{F}'&=4\pi\mu\frac{d^2\mathrm{F}'}{dt^2}, \\ k\nabla^2\mathrm{G}'&=4\pi\mu\frac{d^2\mathrm{G}'}{dt^2}, \\ k\nabla^2\mathrm{H}'&=4\pi\mu\frac{d^2\mathrm{H}'}{dt^2}. \end{aligned}\right\} \quad \ldots \ldots \quad (78)$$

Hence the disturbances indicated by F′, G′, H′ are propagated with the velocity $\mathrm{V}=\sqrt{\frac{k}{4\pi\mu}}$ through the field; and since

$$\frac{d\mathrm{F}'}{dx}+\frac{d\mathrm{G}'}{dy}+\frac{d\mathrm{H}'}{dx}=0,$$

the resultant of these disturbances is in the plane of the wave.

(99) The remaining part of the total disturbances F, G, H being the part depending on χ, is subject to no condition except that expressed in the equation

$$\frac{d\Psi}{dt}+\frac{d^2\chi}{dt^2}=0.$$

If we perform the operation ∇^2 on this equation, it becomes

$$ke=\frac{d\mathrm{J}}{dt}-k\nabla^2\phi(x, y, z). \quad \ldots \ldots \quad (79)$$

Since the medium is a perfect insulator, e, the free electricity, is immoveable, and therefore $\frac{d\mathrm{J}}{dt}$ is a function of x, y, z, and the value of J is either constant or zero, or uniformly increasing or diminishing with the time; so that no disturbance depending on J can be propagated as a wave.

(100) The equations of the electromagnetic field, deduced from purely experimental evidence, show that transversal vibrations only can be propagated. If we were to go beyond our experimental knowledge and to assign a definite density to a substance which

we should call the electric fluid, and select either vitreous or resinous electricity as the representative of that fluid, then we might have normal vibrations propagated with a velocity depending on this density. We have, however, no evidence as to the density of electricity, as we do not even know whether to consider vitreous electricity as a substance or as the absence of a substance.

Hence electromagnetic science leads to exactly the same conclusions as optical science with respect to the direction of the disturbances which can be propagated through the field; both affirm the propagation of transverse vibrations, and both give the same velocity of propagation. On the other hand, both sciences are at a loss when called on to affirm or deny the existence of normal vibrations.

Relation between the Index of Refraction and the Electromagnetic Character of the substance.

(101) The velocity of light in a medium, according to the Undulatory Theory, is

$$\frac{1}{i}V_0,$$

where i is the index of refraction and V_0 is the velocity in vacuum. The velocity, according to the Electromagnetic Theory, is

$$\sqrt{\frac{k}{4\pi\mu}},$$

where, by equations (49) and (71), $k=\frac{1}{D}k_0$, and $k_0=4\pi V_0^2$.

Hence $$D=\frac{i^2}{\mu}, \quad \text{(80)}$$

or the Specific Inductive Capacity is equal to the square of the index of refraction divided by the coefficient of magnetic induction.

Propagation of Electromagnetic Disturbances in a Crystallized Medium.

(102) Let us now calculate the conditions of propagation of a plane wave in a medium for which the values of k and μ are different in different directions. As we do not propose to give a complete investigation of the question in the present imperfect state of the theory as extended to disturbances of short period, we shall assume that the axes of magnetic induction coincide in direction with those of electric elasticity.

(103) Let the values of the magnetic coefficient for the three axes be λ, μ, ν, then the equations of magnetic force (B) become

$$\left.\begin{aligned} \lambda\alpha &= \frac{dH}{dy}-\frac{dG}{dz}, \\ \mu\beta &= \frac{dF}{dz}-\frac{dH}{dx}, \\ \nu\gamma &= \frac{dG}{dx}-\frac{dF}{dy}. \end{aligned}\right\} \quad \text{(81)}$$

The equations of electric currents (C) remain as before.

The equations of electric elasticity (E) will be

$$\left.\begin{aligned} P&=4\pi a^2 f, \\ Q&=4\pi b^2 g, \\ R&=4\pi c^2 h, \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (82)$$

where $4\pi a^2$, $4\pi b^2$, and $4\pi c^2$ are the values of k for the axes of x, y, z.

Combining these equations with (A) and (D), we get equations of the form

$$\frac{1}{\mu\nu}\left(\lambda\frac{d^2F}{dx^2}+\mu\frac{d^2F}{dy^2}+\nu\frac{d^2F}{dz^2}\right)-\frac{1}{\mu\nu}\frac{d}{dx}\left(\lambda\frac{dF}{dx}+\mu\frac{dG}{dy}+\nu\frac{dH}{dz}\right)=\frac{1}{a^2}\left(\frac{d^2F}{dt^2}+\frac{d^2\Psi}{dxdt}\right). \quad . \quad (83)$$

(104) If l, m, n are the direction-cosines of the wave, and V its velocity, and if

$$lx+my+nz-Vt=w, \quad \ldots\ldots\ldots\ldots \quad (84)$$

then F, G, H, and Ψ will be functions of w; and if we put F', G', H', Ψ' for the second differentials of these quantities with respect to w, the equations will be

$$\left.\begin{aligned} &\left(V^2-a^2\left(\frac{m^2}{\nu}+\frac{n^2}{\mu}\right)\right)F'+\frac{a^2lm}{\nu}G'+\frac{a^2ln}{\mu}H'-lV\Psi'=0, \\ &\left(V^2-b^2\left(\frac{n^2}{\lambda}+\frac{l^2}{\nu}\right)\right)G'+\frac{b^2mn}{\lambda}H'+\frac{b^2ml}{\nu}F'-mV\Psi'=0, \\ &\left(V^2-c^2\left(\frac{l^2}{\mu}+\frac{m^2}{\lambda}\right)\right)H'+\frac{c^2nl}{\mu}F'+\frac{c^2nm}{\lambda}G'-nV\Psi'=0. \end{aligned}\right\} \quad \ldots\ldots \quad (85)$$

If we now put

$$\left.\begin{aligned} V^4-V^2\frac{1}{\lambda\mu\nu}&\left\{l^2\lambda(b^2\mu+c^2\nu)+m^2\mu(c^2\nu+a^2\lambda)+n^2\nu(a^2\lambda+b^2\mu)\right\} \\ &+\frac{a^2b^2c^2}{\lambda\mu\nu}\left(\frac{l^2}{a^2}+\frac{m^2}{b^2}+\frac{n^2}{c^2}\right)(l^2\lambda+m^2\mu+n^2\nu)=U, \end{aligned}\right\} \quad \ldots\ldots \quad (86)$$

we shall find

$$F'V^2U-l\Psi'VU=0, \quad \ldots\ldots\ldots\ldots \quad (87)$$

with two similar equations for G' and H'. Hence either

$$V=0, \quad \ldots\ldots\ldots\ldots \quad (88)$$

$$U=0, \quad \ldots\ldots\ldots\ldots \quad (89)$$

or

$$VF'=l\Psi',\ VG'=m\Psi' \text{ and } VH'=n\Psi'. \quad \ldots\ldots \quad (90)$$

The third supposition indicates that the resultant of F', G', H' is in the direction normal to the plane of the wave; but the equations do not indicate that such a disturbance, if possible, could be propagated, as we have no other relation between Ψ' and F', G', H'.

The solution $V=0$ refers to a case in which there is no propagation.

The solution $U=0$ gives two values for V^2 corresponding to values of F', G', H', which are given by the equations

$$\frac{l}{a^2}F'+\frac{m}{b^2}G'+\frac{n}{c^2}H'=0, \quad \ldots \ldots \quad (91)$$

$$\frac{a^2l\lambda}{F'}(b^2\mu-c^2\nu)+\frac{b^2m\mu}{G'}(c^2\nu-a^2\lambda)+\frac{c^2n\nu}{H'}(a^2\lambda-b^2\mu)=0, \quad \ldots \ldots \quad (92)$$

(105) The velocities along the axes are as follows:—

Direction of propagation		x	y	z
Direction of the electric displacements	x		$\frac{a^2}{\nu}$	$\frac{a^2}{\mu}$
	y	$\frac{b^2}{\nu}$		$\frac{b^2}{\lambda}$
	z	$\frac{c^2}{\mu}$	$\frac{c^2}{\lambda}$	

Now we know that in each principal plane of a crystal the ray polarized in that plane obeys the ordinary law of refraction, and therefore its velocity is the same in whatever direction in that plane it is propagated.

If polarized light consists of electromagnetic disturbances in which the electric displacement is in the plane of polarization, then

$$a^2=b^2=c^2. \quad \ldots \ldots \quad (93)$$

If, on the contrary, the electric displacements are perpendicular to the plane of polarization,

$$\lambda=\mu=\nu. \quad \ldots \ldots \quad (94)$$

We know, from the magnetic experiments of FARADAY, PLÜCKER, &c., that in many crystals λ, μ, ν are unequal.

The experiments of KNOBLAUCH* on electric induction through crystals seem to show that a, b and c, may be different.

The inequality, however, of λ, μ, ν is so small that great magnetic forces are required to indicate their difference, and the differences do not seem of sufficient magnitude to account for the double refraction of the crystals.

On the other hand, experiments on electric induction are liable to error on account of minute flaws, or portions of conducting matter in the crystal.

Further experiments on the magnetic and dielectric properties of crystals are required before we can decide whether the relation of these bodies to magnetic and electric forces is the same, when these forces are permanent as when they are alternating with the rapidity of the vibrations of light.

* Philosophical Magazine, 1852.

Relation between Electric Resistance and Transparency.

(106) If the medium, instead of being a perfect insulator, is a conductor whose resistance per unit of volume is ϱ, then there will be not only electric displacements, but true currents of conduction in which electrical energy is transformed into heat, and the undulation is thereby weakened. To determine the coefficient of absorption, let us investigate the propagation along the axis of x of the transverse disturbance G.

By the former equations

$$\frac{d^2G}{dx^2}=-4\pi\mu(q')$$

$$=-4\pi\mu\left(\frac{df}{dt}+q\right) \text{ by (A)},$$

$$\frac{d^2G}{dx^2}=+4\pi\mu\left(\frac{1}{k}\frac{d^2G}{dt^2}-\frac{1}{\varrho}\frac{dG}{dt}\right) \text{ by (E) and (F)}. \quad (95)$$

If G is of the form

$$G=e^{-px}\cos(qx+nt), \quad (96)$$

we find that

$$p=\frac{2\pi\mu}{\varrho}\frac{n}{q}=\frac{2\pi\mu}{\varrho}\frac{V}{i}, \quad (97)$$

where V is the velocity of light in air, and i is the index of refraction. The proportion of incident light transmitted through the thickness x is

$$e^{-2px}. \quad (98)$$

Let R be the resistance in electromagnetic measure of a plate of the substance whose thickness is x, breadth b, and length l, then

$$R=\frac{l\varrho}{bx},$$

$$2px=4\pi\mu\frac{V}{i}\frac{l}{bR}. \quad (99)$$

(107) Most transparent solid bodies are good insulators, whereas all good conductors are very opaque.

Electrolytes allow a current to pass easily and yet are often very transparent. We may suppose, however, that in the rapidly alternating vibrations of light, the electromotive forces act for so short a time that they are unable to effect a complete separation between the particles in combination, so that when the force is reversed the particles oscillate into their former position without loss of energy.

Gold, silver, and platinum are good conductors, and yet when reduced to sufficiently thin plates they allow light to pass through them. If the resistance of gold is the same for electromotive forces of short period as for those with which we make experiments, the amount of light which passes through a piece of gold-leaf, of which the resistance was determined by Mr. C. HOCKIN, would be only 10^{-50} of the incident light, a totally imperceptible quantity. I find that between $\frac{1}{500}$ and $\frac{1}{1000}$ of green light gets through such gold-leaf. Much of this is transmitted through holes and cracks; there is enough, however, transmitted through the gold itself to give a strong green hue to the transmitted light. This result cannot be reconciled with the electromagnetic theory of light, unless we suppose that there is less loss of energy when the electromotive forces are reversed with the rapidity of the vibrations of light than when they act for sensible times, as in our experiments.

Absolute Values of the Electromotive and Magnetic Forces called into play in the Propagation of Light.

(108) If the equation of propagation of light is

$$F = A\cos\frac{2\pi}{\lambda}(z - Vt),$$

the electromotive force will be

$$P = -A\frac{2\pi}{\lambda}V\sin\frac{2\pi}{\lambda}(z - Vt);$$

and the energy per unit of volume will be

$$\frac{P^2}{8\pi\mu V^2},$$

where P represents the greatest value of the electromotive force. Half of this consists of magnetic and half of electric energy.

The energy passing through a unit of area is

$$W = \frac{P^2}{8\pi\mu V};$$

so that

$$P = \sqrt{8\pi\mu VW},$$

where V is the velocity of light, and W is the energy communicated to unit of area by the light in a second.

According to POUILLET'S data, as calculated by Professor W. THOMSON *, the mechanical value of direct sunlight at the Earth is

83·4 foot-pounds per second per square foot.

This gives the maximum value of P in direct sunlight at the Earth's distance from the Sun,

$$P = 60,000,000,$$

or about 600 DANIELL'S cells per metre.

At the Sun's surface the value of P would be about

13,000 DANIELL'S cells per metre.

At the Earth the maximum magnetic force would be ·193 †.

At the Sun it would be 4·13.

These electromotive and magnetic forces must be conceived to be reversed twice in every vibration of light; that is, more than a thousand million million times in a second.

* Transactions of the Royal Society of Edinburgh, 1854 ("Mechanical Energies of the Solar System").

† The horizontal magnetic force at Kew is about 1·76 in metrical units.

PART VII.—CALCULATION OF THE COEFFICIENTS OF ELECTROMAGNETIC INDUCTION.

General Methods.

(109) The electromagnetic relations between two conducting circuits, A and B, depend upon a function M of their form and relative position, as has been already shown.

M may be calculated in several different ways, which must of course all lead to the same result.

First Method. M is the electromagnetic momentum of the circuit B when A carries a unit current, or

$$\mathrm{M}=\int\left(\mathrm{F}\frac{dx}{ds'}+\mathrm{G}\frac{dy}{ds'}+\mathrm{H}\frac{dz}{ds'}\right)ds',$$

where F, G, H are the components of electromagnetic momentnm due to a unit current in A, and ds' is an element of length of B, and the integration is performed round the circuit of B.

To find F, G, H, we observe that by (B) and (C)

$$\frac{d^2\mathrm{F}}{dx^2}+\frac{d^2\mathrm{F}}{dy^2}+\frac{d^2\mathrm{F}}{dz^2}=-4\pi\mu p',$$

with corresponding equations for G and H, p', q', and r' being the components of the current in A.

Now if we consider only a single element ds of A, we shall have

$$p'=\frac{dx}{ds}ds, \qquad q'=\frac{dy}{ds}ds, \qquad r'=\frac{dz}{ds}ds,$$

and the solution of the equation gives

$$\mathrm{F}=\frac{\mu}{\varrho}\frac{dx}{ds}ds, \qquad \mathrm{G}=\frac{\mu}{\varrho}\frac{dy}{ds}ds, \qquad \mathrm{H}=\frac{\mu}{\varrho}\frac{dz}{ds}ds,$$

where ϱ is the distance of any point from ds. Hence

$$\mathrm{M}=\iint\frac{\mu}{\varrho}\left(\frac{dx}{ds}\frac{dx}{ds'}+\frac{dy}{ds}\frac{dy}{ds'}+\frac{dz}{ds}\frac{dz}{ds'}\right)dsds'$$

$$=\iint\frac{\mu}{\varrho}\cos\theta dsds',$$

where θ is the angle between the directions of the two elements ds, ds', and ϱ is the distance between them, and the integration is performed round both circuits.

In this method we confine our attention during integration to the two linear circuits alone.

(110) Second Method. M is the number of lines of magnetic force which pass through the circuit B when A carries a unit current, or

$$\mathrm{M}=\Sigma(\mu\alpha l+\mu\beta m+\mu\gamma n)d\mathrm{S}',$$

where $\mu\alpha$, $\mu\beta$, $\mu\gamma$ are the components of magnetic induction due to unit current in A, S′ is a surface bounded by the current B, and l, m, n are the direction-cosines of the normal to the surface, the integration being extended over the surface.

We may express this in the form

$$M=\mu\Sigma\frac{1}{\varrho^2}\sin\theta\sin\theta'\sin\phi dS'ds,$$

where dS' is an element of the surface bounded by B, ds is an element of the circuit A, ϱ is the distance between them, θ and θ' are the angles between ϱ and ds and between ϱ and the normal to dS' respectively, and ϕ is the angle between the planes in which θ and θ' are measured. The integration is performed round the circuit A and over the surface bounded by B.

This method is most convenient in the case of circuits lying in one plane, in which case $\sin\theta=1$, and $\sin\phi=1$.

111. Third Method. M is that part of the intrinsic magnetic energy of the whole field which depends on the product of the currents in the two circuits, each current being unity.

Let α, β, γ be the components of magnetic intensity at any point due to the first circuit, α', β', γ' the same for the second circuit; then the intrinsic energy of the element of volume dV of the field is

$$\frac{\mu}{8\pi}\left((\alpha+\alpha')^2+(\beta+\beta')^2+(\gamma+\gamma')^2\right)dV.$$

The part which depends on the product of the currents is

$$\frac{\mu}{4\pi}(\alpha\alpha'+\beta\beta'+\gamma\gamma')dV.$$

Hence if we know the magnetic intensities I and I′ due to unit current in each circuit, we may obtain M by integrating

$$\frac{\mu}{4\pi}\Sigma\mu I\,I'\cos\theta dV$$

over all space, where θ is the angle between the directions of I and I′.

Application to a Coil.

(112) To find the coefficient (M) of mutual induction between two circular linear conductors in parallel planes, the distance between the curves being everywhere the same, and small compared with the radius of either.

If r be the distance between the curves, and a the radius of either, then when r is very small compared with a, we find by the second method, as a first approximation,

$$M=4\pi a\left(\log_e\frac{8a}{r}-2\right).$$

To approximate more closely to the value of M, let a and a_1 be the radii of the circles, and b the distance between their planes; then

$$r^2=(a-a_1)^2+b^2.$$

We obtain M by considering the following conditions:—

1st. M must fulfil the differential equation

$$\frac{d^2\mathrm{M}}{da^2}+\frac{d^2\mathrm{M}}{db^2}+\frac{1}{a}\frac{d\mathrm{M}}{da}=0.$$

This equation being true for any magnetic field symmetrical with respect to the common axis of the circles, cannot of itself lead to the determination of M as a function of a, a_1, and b. We therefore make use of other conditions.

2ndly. The value of M must remain the same when a and a_1 are exchanged.

3rdly. The first two terms of M must be the same as those given above.

M may thus be expanded in the following series:—

$$\mathrm{M}=4\pi a\log\frac{8a}{r}\left\{1+\frac{1}{2}\frac{a-a_1}{a}+\frac{1}{16}\frac{3b^2+(a_1-a)^2}{a^2}-\frac{1}{32}\frac{(3b^2+(a-a_1)^2)(a-a_1)}{a^3}+\&c.\right\}$$
$$-4\pi a\left\{2+\frac{1}{2}\frac{a-a_1}{a}+\frac{1}{16}\frac{b^2-3(a-a_1^2)}{a^2}-\frac{1}{48}\frac{(6b^2-(a-a_1)^2)(a-a_1)}{a^3}+\&c.\right\}.$$

(113) We may apply this result to find the coefficient of self-induction (L) of a circular coil of wire whose section is small compared with the radius of the circle.

Let the section of the coil be a rectangle, the breadth in the plane of the circle being c, and the depth perpendicular to the plane of the circle being b.

Let the mean radius of the coil be a, and the number of windings n; then we find, by integrating,

$$\mathrm{L}=\frac{n^2}{b^2c^2}\iiiint\mathrm{M}(xy\,x'y')dx\,dy\,dx'\,dy',$$

where $\mathrm{M}(xy\,x'y')$ means the value of M for the two windings whose coordinates are xy and $x'y'$ respectively; and the integration is performed first with respect to x and y over the rectangular section, and then with respect to x' and y' over the same space.

$$\mathrm{L}=4\pi n^2a\left\{\log_e\frac{8a}{r}+\frac{1}{12}-\frac{4}{3}\left(\theta-\frac{\pi}{4}\right)\cot 2\theta-\frac{\pi}{3}\cos 2\theta-\frac{1}{6}\cot^2\theta\log\cos\theta-\frac{1}{6}\tan^2\theta\log\sin\theta\right\}$$
$$+\frac{\pi n^2r^2}{24a}\left\{\log\frac{8a}{r}(2\sin^2\theta+1)+3\cdot45+27\cdot475\cos^2\theta-3\cdot2\left(\frac{\pi}{2}-\theta\right)\frac{\sin^3\theta}{\cos\theta}+\frac{1}{5}\frac{\cos^4\theta}{\sin^2\theta}\log\cos\theta\right.$$
$$\left.+\frac{13}{3}\frac{\sin^4\theta}{\cos^2\theta}\log\sin\theta\right\}+\&c.$$

Here $a=$ mean radius of the coil.

" $r=$ diagonal of the rectangular section $=\sqrt{b^2+c^2}$.

" $\theta=$ angle between r and the plane of the circle.

" $n=$ number of windings.

The logarithms are Napierian, and the angles are in circular measure.

In the experiments made by the Committee of the British Association for determining a standard of Electrical Resistance, a double coil was used, consisting of two nearly equal coils of rectangular section, placed parallel to each other, with a small interval between them.

The value of L for this coil was found in the following way.

The value of L was calculated by the preceding formula for six different cases, in which the rectangular section considered has always the same breadth, while the depth was

$$A, \; B, \; C, \quad A+B, \quad B+C, \quad A+B+C,$$

and $n=1$ in each case.

Calling the results

$$L(A), \quad L(B), \quad L(C), \; \&c.,$$

we calculate the coefficient of mutual induction M(AC) of the two coils thus,

$$2ACM(AC)=(A+B+C)^2L(A+B+C)-(A+B)^2L(A+B)-(B+C)^2L(B+C)+B^2L(B).$$

Then if n_1 is the number of windings in the coil A and n_2 in the coil B, the coefficient of self-induction of the two coils together is

$$L=n_1^2L(A)+2n_1n_2L(AC)+n_2^2L(B).$$

(114) These values of L are calculated on the supposition that the windings of the wire are evenly distributed so as to fill up exactly the whole section. This, however, is not the case, as the wire is generally circular and covered with insulating material. Hence the current in the wire is more concentrated than it would have been if it had been distributed uniformly over the section, and the currents in the neighbouring wires do not act on it exactly as such a uniform current would do.

The corrections arising from these considerations may be expressed as numerical quantities, by which we must multiply the length of the wire, and they are the same whatever be the form of the coil.

Let the distance between each wire and the next, on the supposition that they are arranged in square order, be D, and let the diameter of the wire be d, then the correction for diameter of wire is

$$+2\left(\log\frac{D}{d}+\frac{4}{3}\log 2+\frac{\pi}{3}-\frac{11}{6}\right).$$

The correction for the eight nearest wires is

$$+0{\cdot}0236.$$

For the sixteen in the next row

$$+0{\cdot}00083.$$

These corrections being multiplied by the length of wire and added to the former result, give the true value of L, considered as the measure of the potential of the coil on itself for unit current in the wire when that current has been established for some time, and is uniformly distributed through the section of the wire.

(115) But at the commencement of a current and during its variation the current is not uniform throughout the section of the wire, because the inductive action between different portions of the current tends to make the current stronger at one part of the section than at another. When a uniform electromotive force P arising from any cause

acts on a cylindrical wire of specific resistance ϱ, we have

$$p\varrho = \mathrm{P} - \frac{d\mathrm{F}}{dt},$$

where F is got from the equation

$$\frac{d^2\mathrm{F}}{dr^2} + \frac{1}{r}\frac{d\mathrm{F}}{dr} = -4\pi\mu p,$$

r being the distance from the axis of the cylinder.

Let one term of the value of F be of the form $\mathrm{T}r^n$, where T is a function of the time, then the term of p which produced it is of the form

$$-\frac{1}{4\pi\mu} n^2 \mathrm{T} r^{n-2}.$$

Hence if we write

$$\mathrm{F} = \mathrm{T} + \frac{\mu\pi}{\varrho}\left(-\mathrm{P} + \frac{d\mathrm{T}}{dt}\right) r^2 + \overline{\frac{\mu\pi}{\varrho}}\Big|^2 \frac{1}{1^2.2^2} \frac{d\mathrm{T}^2}{dt^2} r^4 + \text{\&c.}$$

$$p\varrho = \left(\mathrm{P} + \frac{d\mathrm{T}}{dt}\right) - \frac{\mu\pi}{\varrho}\frac{d^2\mathrm{T}}{dt^2} r^2 - \overline{\frac{\mu\pi}{\varrho}}\Big|^2 \frac{1}{1^2.2^2}\frac{d^3\mathrm{T}}{dt^3} r^4 - \text{\&c.}$$

The total counter current of self-induction at any point is

$$\int\left(\frac{\mathrm{P}}{\varrho} - p\right) dt = \frac{1}{\varrho}\mathrm{T} + \frac{\mu\pi}{\varrho^2}\frac{d\mathrm{T}}{dt} r^2 + \frac{\mu^2\pi^2}{\varrho^3}\frac{1}{1^2 2^2}\frac{d^2\mathrm{T}}{dt^2} r^4 + \text{\&c.}$$

from $t=0$ to $t=\infty$.

When $t=0$, $p=0$, $\therefore \left(\frac{d\mathrm{T}}{dt}\right)_0 = \mathrm{P}$, $\left(\frac{d^2\mathrm{T}}{dt^2}\right)_0 = 0$, &c.

When $t=\infty$, $p=\frac{\mathrm{P}}{\varrho}$, $\therefore \left(\frac{d\mathrm{T}}{dt}\right)_\infty = 0$, $\left(\frac{d^2\mathrm{T}}{dt^2}\right)_\infty = 0$, &c.

$$\int_0^\infty \int_0^r 2\pi\left(\frac{\mathrm{P}}{\varrho} - p\right) r\,dr\,dt = \frac{1}{\varrho}\mathrm{T}\pi r^2 + \frac{1}{2}\frac{\mu\pi^2}{\varrho^2}\frac{d\mathrm{T}}{dt} r^4 + \frac{\mu^2\pi^3}{\varrho^3}\frac{1}{1^2.2^2.3}\frac{d^2\mathrm{T}}{dt^2} r^6 + \text{\&c.}$$

from t=0 to $=\infty$.

When $t=0$, $p=0$ throughout the section, $\therefore \left(\frac{d\mathrm{T}}{dt}\right)_0 = \mathrm{P}$, $\left(\frac{d^2\mathrm{T}}{dt^2}\right)_0 = 0$, &c.

When $t=\infty$, $p=0$ throughout . . . . . $\therefore \left(\frac{d\mathrm{T}}{dt}\right)_\infty = 0$, $\left(\frac{d^2\mathrm{T}}{dt^2}\right)_\infty = 0$, &c.

Also if l be the length of the wire, and R its resistance,

$$\mathrm{R} = \frac{\varrho l}{\pi r^2};$$

and if C be the current when established in the wire, $\mathrm{C} = \frac{\mathrm{P}l}{\mathrm{R}}$.

The total counter current may be written

$$\frac{l}{\mathrm{R}}(\mathrm{T}_\infty - \mathrm{T}_0) - \frac{1}{2}\mu\frac{l}{\mathrm{R}}\mathrm{C} = -\frac{\mathrm{LC}}{\mathrm{R}} \text{ by § (35).}$$

Now if the current instead of being variable from the centre to the circumference of the section of the wire had been the same throughout, the value of F would have been

$$F=T+\mu\gamma\left(1-\frac{r^2}{r_0^2}\right),$$

where γ is the current in the wire at any instant, and the total countercurrent would have been

$$\int_0^\infty\int_0^r \frac{1}{\varrho}\frac{dF}{dt}2\pi r dr=\frac{l}{R}(T_\infty-T_0)-\frac{3}{4}\mu\frac{l}{R}C=-\frac{L'C}{R}, \text{ say.}$$

Hence

$$L=L'-\tfrac{1}{4}\mu l,$$

or the value of L which must be used in calculating the self-induction of a wire for variable currents is less than that which is deduced from the supposition of the current being constant throughout the section of the wire by $\frac{1}{4}\mu l$, where l is the length of the wire, and μ is the coefficient of magnetic induction for the substance of the wire.

(116) The dimensions of the coil used by the Committee of the British Association in their experiments at King's College in 1864 were as follows:—

	metre.
Mean radius	$=a=\cdot158194$
Depth of each coil	$=b=\cdot01608$
Breadth of each coil	$=c=\cdot01841$
Distance between the coils. . .	$=\cdot02010$
Number of windings	$n=313$
Diameter of wire	$=\cdot00126$

The value of L derived from the first term of the expression is 437440 metres.

The correction depending on the radius not being infinitely great compared with the section of the coil as found from the second term is −7345 metres.

The correction depending on the diameter of the wire is per unit of length	+·44997
Correction of eight neighbouring wires.	+·0236
For sixteen wires next to these	+·0008
Correction for variation of current in different parts of section	−·2500
Total correction per unit of length	·22437
Length	311·236 metres.
Sum of corrections of this kind	70 „
Final value of L by calculation	430165 „

This value of L was employed in reducing the observations, according to the method explained in the Report of the Committee*. The correction depending on L varies as the square of the velocity. The results of sixteen experiments to which this correction had been applied, and in which the velocity varied from 100 revolutions in seventeen seconds to 100 in seventy-seven seconds, were compared by the method of

* British Association Reports, 1863, p. 169.

least squares to determine what further correction depending on the square of the velocity should be applied to make the outstanding errors a minimum.

The result of this examination showed that the calculated value of L should be multiplied by 1·0618 to obtain the value of L, which would give the most consistent results.

We have therefore L by calculation	430165	metres.
Probable value of L by method of least squares	456748	„
Result of rough experiment with the Electric Balance (see § 46)	410000	„

The value of L calculated from the dimensions of the coil is probably much more accurate than either of the other determinations.

CHAPTER V

ON WORK AND ENERGY

72. DEFINITIONS

WORK *is the act of* producing a change of configuration in a system in opposition to a force which resists that change.*

ENERGY *is the capacity of doing work.*

When the nature of a material system is such that if, after the system has undergone any series of changes it is brought back in any manner to its original state, the whole work done by external agents on the system is equal to the whole work done by the system in overcoming external forces, the system is called a CONSERVATIVE SYSTEM†.

73. PRINCIPLE OF CONSERVATION OF ENERGY

The progress of physical science has led to the discovery and investigation of different forms of energy, and to the establishment of the doctrine that all material systems may be regarded as conservative systems, *provided* that all the different forms of energy which exist in these systems are taken into account.

This doctrine, considered as a deduction from observation and experiment, can, of course, assert no more than that no instance of a non-conservative system has hitherto been discovered.

As a scientific or science-producing doctrine, how-

* The work done is a quantitative measure of the effort expended in deranging the system, in terms of the consumption of energy that is required to give effect to it.

The idea of work implies a fund of energy, from which the work is supplied.

† As distinguished from a system in which the energy available for work becomes gradually degraded to less available forms by frictional agencies, called a *Dissipative System*. Cf. Art. 93.

ever, it is always acquiring additional credibility from the constantly increasing number of deductions which have been drawn from it, and which are found in all cases to be verified by experiment.

In fact the doctrine of the Conservation of Energy is the one generalised statement which is found to be consistent with fact, not in one physical science only, but in all.

When once apprehended it furnishes to the physical inquirer a principle on which he may hang every known law relating to physical actions, and by which he may be put in the way to discover the relations of such actions in new branches of science*.

For such reasons the doctrine is commonly called the Principle of the Conservation of Energy.

74. General Statement of the Principle of the Conservation of Energy

The total energy of any material system is a quantity which can neither be increased nor diminished by any action between the parts of the system, though it may be transformed into any of the forms of which energy is susceptible.

If, by the action of some agent external to the system, the configuration of the system is changed, while the forces of the system resist this change of configuration, the external agent is said to do work on the system. In this case the energy of the system is increased by the amount of work done on it by the external agent.

If, on the contrary, the forces of the system produce a change of configuration which is resisted by the external agent, the system is said to do work on the

* Every law relating to the forces of statical or steady systems is involved implicitly in the complete expression for the Energy of the system. But in a kinetic system, where force is being used in producing energy of motion, a more elaborate principle is required, that of Least Action, for example. See *infra*, Chapter IX.

external agent, and the energy of the system is diminished by the amount of work which it does.

Work, therefore, is a transference of energy from one system to another; the system which gives out energy is said to do work on the system which receives it, and the amount of energy given out by the first system is always exactly equal to that received by the second.

If, therefore, we include both systems in one larger system, the energy of the total system is neither increased nor diminished by the action of the one partial system on the other.

75. Measurement of Work

Work done by an external agent on a material system may be described as a change* in the configuration of the system taking place under the action of an external force tending to produce that change.

Thus, if one pound is lifted one foot from the ground by a man in opposition to the force of gravity, a certain amount of work is done by the man, and this quantity is known among engineers as one foot-pound.

Here the man is the external agent, the material system consists of the earth and the pound, the change of configuration is the increase of the distance between the matter of the earth and the matter of the pound, and the force is the upward force exerted by the man in lifting the pound, which is equal and opposite to the weight of the pound. To raise the pound a foot higher would, if gravity were a uniform force, require exactly the same amount of work. It is true that gravity is not really uniform, but diminishes as we ascend from the earth's surface, so that a foot-pound is not an accurately

* See footnote, Art. 72.

These ideas, leading to an estimate of the total effect by *work* done rather than *momentum* produced, are of the kind that were enforced by Leibniz. What was then mainly needed to avoid confusion was a set of names for the different effects.

known quantity, unless we specify the intensity of gravity at the place. But for the purpose of illustration we may assume that gravity is uniform for a few feet of ascent, and in that case the work done in lifting a pound would be one foot-pound for every foot the pound is lifted.

To raise twenty pounds of water ten feet high requires 200 foot-pounds of work. To raise one pound ten feet high requires ten foot-pounds, and as there are twenty pounds the whole work is twenty times as much, or two hundred foot-pounds.

The quantity of work done is, therefore, proportional to the product of the numbers representing the force exerted and the displacement in the direction of the force.

In the case of a foot-pound the force is the weight of a pound—a quantity which, as we know, is different in different places. The weight of a pound expressed in absolute measure is numerically equal to the intensity of gravity, the quantity denoted by g, the value of which in poundals to the pound varies from 32·227 at the poles to 32·117 at the equator, and diminishes without limit as we recede from the earth. In dynes to the gramme it varies from 978·1 to 983·1. Hence, in order to express work in a uniform and consistent manner, we must multiply the number of foot-pounds by the number representing the intensity of gravity at the place. The work is thus reduced to foot-poundals. We shall always understand work to be measured in this manner and reckoned in foot-poundals when no other system of measurement is mentioned. When work is expressed in foot-pounds the system is that of *gravitation-measures*, which is not a complete system unless we also know the intensity of gravity at the place.

In the metrical system the unit of work is the Erg, which is the work done by a dyne acting through a centimetre. There are 421393·8 ergs in a foot-poundal.

76. Potential Energy

The work done by a man in raising a heavy body is done in overcoming the attraction between the earth and that body. The energy of the material system, consisting of the earth and the heavy body, is thereby increased. If the heavy body is the leaden weight of a clock, the energy of the clock is increased by winding it up, so that the clock is able to go for a week in spite of the friction of the wheels and the resistance of the air to the motion of the pendulum, and also to give out energy in other forms, such as the communication of the vibrations to the air, by which we hear the ticking of the clock.

When a man winds up a watch he does work in changing the form of the mainspring by coiling it up. The energy of the mainspring is thereby increased, so that as it uncoils itself it is able to keep the watch going.

In both these cases the energy communicated to the system depends upon a change of configuration.

77. Kinetic Energy

But in a very important class of phenomena the work is done in changing the velocity of the body on which it acts. Let us take as a simple case that of a body moving without rotation under the action of a force. Let the mass of the body be M pounds, and let a force of F poundals act on it in the line of motion during an interval of time, T seconds. Let the velocity at the beginning of the interval be V and that at the end V' feet per second, and let the distance travelled by the body during the time be S feet. The original momentum is MV, and the final momentum is MV', so that the increase of momentum is $M(V' - V)$, and this, by the second law of motion, is equal to FT, the *impulse* of the force F acting for the time T. Hence

$$FT = M(V' - V) \qquad \ldots\ldots(1).$$

Since the velocity increases uniformly with the time [when the force is constant], the mean velocity is the arithmetical mean of the original and final velocities, or $\frac{1}{2}(V' + V)$.

We can also determine the mean velocity by dividing the space S by the time T, during which it is described. Hence

$$\frac{S}{T} = \frac{1}{2}(V' + V) \qquad \text{......(2).}$$

Multiplying the corresponding members of equations (1) and (2) each by each we obtain

$$FS = \tfrac{1}{2}MV'^2 - \tfrac{1}{2}MV^2 \qquad \text{......(3).}$$

Here FS is the work done by the force F acting on the body while it moves through the space S in the direction of the force, and this is equal to the excess of $\frac{1}{2}MV'^2$ above $\frac{1}{2}MV^2$. If we call $\frac{1}{2}MV^2$, or half the product of the mass into the square of the velocity, the *kinetic energy* of the body at first, then $\frac{1}{2}MV'^2$ will be the kinetic energy after the action of the force F through the space S. The energy is here expressed in foot-poundals.

We may now express the equation in words by saying that the work done by the force F in changing the motion of the body is measured by the increase of the kinetic energy of the body during the time that the force acts.

We have proved that this is true, when the interval of time is so small that we may consider the force as constant during that time, and the mean velocity during the interval as the arithmetical mean of the velocities at the beginning and end of the interval. This assumption, which is exactly true when the force is constant, however long the interval may be, becomes in every case more and more nearly true as the interval of time taken becomes smaller and smaller. By dividing the whole time of action into small parts, and proving that in each of these the work done is equal to the increase of the kinetic energy of the body, we may, by adding the successive portions of the work and the successive increments of energy, arrive at the result that the total work done by the force is equal to the total increase of kinetic energy.

If the force acts on the body in the direction opposite to its motion, the kinetic energy of the body will be diminished instead of being increased, and the force, instead of doing work on the body, will act as a resistance, which the body, in its motion, overcomes. Hence a moving body, as long as it is in motion, can do work in overcoming resistance, and the work done by the moving body is equal to the diminution of its kinetic energy, till at last, when the body is brought to rest, its kinetic energy is exhausted, and the whole work it has done is then equal to the whole kinetic energy which it had at first.

We now see the appropriateness of the name *kinetic energy*, which we have hitherto used merely as a name to denote the product $\frac{1}{2}MV^2$. For the energy of a body has been defined as the capacity which it has of doing work, and it is measured by the work which it can do. The *kinetic* energy of a body is the energy it has in virtue of being in *motion*, and we have now shown that its value is expressed by $\frac{1}{2}MV^2$ or $\frac{1}{2}MV \times V$, that is, half the product of its momentum into its velocity.

78. Oblique Forces

If the force acts on the body at right angles to the direction of its motion it does no work on the body, and it alters the direction but not the magnitude of the velocity. The kinetic energy, therefore, which depends on the square of the velocity, remains unchanged.

If the direction of the force is neither coincident with, nor at right angles to, that of the motion of the body we may resolve the force into two components, one of which is at right angles to the direction of motion, while the other is in the direction of motion (or in the opposite direction).

The first of these components may be left out of consideration in all calculations about energy, since it neither does work on the body nor alters its kinetic energy.

The second component is that which we have already considered. When it is in the direction of motion it increases the kinetic energy of the body by the amount of work which it does on the body. When it is in the opposite direction the kinetic energy of the body is diminished by the amount of work which the body does against the force.

Hence in all cases the increase of kinetic energy is equal to the work done on the body by external agency, and the diminution of kinetic energy is equal to the work done by the body against external resistance.

79. Kinetic Energy of Two Particles referred to their Centre of Mass

The kinetic energy of a material system is equal to the kinetic energy of a mass equal to that of the system moving with the velocity of the centre of mass of the system, together with the kinetic energy due to the motion of the parts of the system relative to its centre of mass.

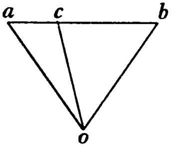


Fig. 10.

Let us begin with the case of two particles whose masses are A and B, and whose velocities are represented in the diagram of velocities by the lines *oa* and *ob*. If c is the centre of mass of a particle equal to A placed at a, and a particle equal to B placed at b, then *oc* will represent the velocity of the centre of mass of the two particles.

The kinetic energy of the system is the sum of the kinetic energies of the particles, or

$$T = \tfrac{1}{2}Aoa^2 + \tfrac{1}{2}Bob^2.$$

Expressing oa^2 and ob^2 in terms of oc, ca and cb, and the angle $oca = \theta$,

$$T = \tfrac{1}{2}Aoc^2 + \tfrac{1}{2}Aca^2 - Aoc.ca\cos\theta$$
$$+ \tfrac{1}{2}Boc^2 + \tfrac{1}{2}Bcb^2 - Boc.cb\cos\theta.$$

But since c is the centre of mass of A at a, and B at b,

$$Aca + Bcb = 0.$$

Hence adding

$$T = \tfrac{1}{2}(A+B)\,oc^2 + \tfrac{1}{2}Aca^2 + \tfrac{1}{2}Bcb^2,$$

or, the kinetic energy of the system of two particles A and B is equal to that of a mass equal to $(A+B)$ moving with the velocity of the centre of mass, together with that of the motion of the particles relative to the centre of mass.

80. Kinetic Energy of a Material System referred to its Centre of Mass

We have begun with the case of two particles, because the motion of a particle is assumed to be that of its centre of mass, and we have proved our proposition true for a system of two particles. But if the proposition is true for each of two material systems taken separately, it must be true of the system which they form together. For if we now suppose oa and ob to represent the velocities of the centres of mass of two material systems A and B, then oc will represent the velocity of the centre of mass of the combined system $A+B$, and if T_A represents the kinetic energy of the motion of the system A relative to its own centre of mass, and T_B the same for the system B, then if the proposition is true for the systems A and B taken separately, the kinetic energy of A is

$$\tfrac{1}{2}Aoa^2 + T_A,$$

and that of B $\qquad \tfrac{1}{2}Bob^2 + T_B.$

The kinetic energy of the whole is therefore

$$\tfrac{1}{2}Aoa^2 + \tfrac{1}{2}Bob^2 + T_A + T_B,$$

or, $$\tfrac{1}{2}(A + B)\,oc^2 + \tfrac{1}{2}Aca^2 + T_A + \tfrac{1}{2}Bcb^2 + T_B.$$

The first term represents the kinetic energy of a mass equal to that of the whole system moving with the velocity of the centre of mass of the whole system.

The second and third terms, taken together, represent the kinetic energy of the system A relative to the centre of gravity of the whole system, and the fourth and fifth terms represent the same for the system B.

Hence if the proposition is true for the two systems A and B taken separately, it is true for the system compounded of A and B. But we have proved it true for the case of two particles; it is therefore true for three, four, or any other number of particles, and therefore for any material system.

The kinetic energy of a system referred to its centre of mass is less than its kinetic energy when referred to any other point.

For the latter quantity exceeds the former by a quantity equal to the kinetic energy of a mass equal to that of the whole system moving with the velocity of the centre of mass relative to the other point, and since all kinetic energy is essentially positive, this excess must be positive.

81. Available Kinetic Energy

We have already seen in Article 64 that the mutual action between the parts of a material system cannot change the velocity of the centre of mass of the system. Hence that part of the kinetic energy of the system which depends on the motion of the centre of mass cannot be affected by any action internal to the system. It is therefore impossible, by means of the mutual action of the parts of the system, to convert this part of the energy into work. As far as the system itself is concerned, this energy is unavailable. It can be converted into work only by means of the action between this system and some other material system external to it.

Hence if we consider a material system unconnected with any other system, its available kinetic energy is that which is due to the motions of the parts of the system relative to its centre of mass.

Let us suppose that the action between the parts of the system is such that after a certain time the configuration of the system becomes invariable, and let us call this process the solidification of the system. We have shown that the angular momentum of the whole system is not changed by any mutual action of its parts. Hence if the original angular momentum is zero, the system, when its form becomes invariable, will not rotate about its centre of mass, but if it moves at all will move parallel to itself, and the parts will be at rest relative to the centre of mass. In this case therefore the whole available energy will be converted into work by the mutual action of the parts during the solidification of the system.

If the system has angular momentum, it will have the same angular momentum when solidified. It will therefore rotate about its centre of mass, and will therefore still have energy of motion relative to its centre of mass, and this remaining kinetic energy has not been converted into work.

But if the parts of the system are allowed to separate from one another in directions perpendicular to the axis of the angular momentum of the system, and if the system when thus expanded is solidified, the remaining kinetic energy of rotation round the centre of mass will be less and less the greater the expansion of the system, so that by sufficiently expanding the system [before it is solidified] we may make the remaining kinetic energy as small as we please, so that the whole kinetic energy relative to the centre of mass of the system may be converted into work within the system.

82. Potential Energy

The potential energy of a material system is the capacity which it has of doing work [on other systems] depending on other circumstances than the motion of the system. In other words, potential energy is that energy which is not kinetic.

In the theoretical material system which we build up in our imagination from the fundamental ideas of matter and motion, there are no other conditions present except the configuration and motion of the different masses of which the system is composed. Hence in such a system the circumstances upon which the energy must depend are motion and configuration only, so that, as the kinetic energy depends on the motion, the potential energy must depend on the configuration.

In many real material systems we know that part of the energy does depend on the configuration. Thus the mainspring of a watch has more energy when coiled up than when partially uncoiled, and two bar magnets have more energy when placed side by side with their similar poles turned the same way than when their dissimilar poles are placed next each other.

83. Elasticity

In the case of the spring we may trace the connexion between the coiling of the spring and the force which it exerts somewhat further by conceiving the spring divided (in imagination) into very small parts or elements. When the spring is coiled up, the form of each of these small parts is altered, and such an alteration of the form of a solid body is called a Strain.

In solid bodies strain is accompanied with internal force or stress; those bodies in which the stress depends simply on the strain are called Elastic, and the property of exerting stress when strained is called Elasticity.

We thus find that the coiling of the spring involves the strain of its elements, and that the external force which the spring exerts is the resultant of the stresses in its elements.

We thus substitute for the immediate relation between the coiling of the spring and the force which it exerts, a relation between the strains and stresses of the elements of the spring; that is to say, for a single displacement and a single force, the relation between which may in some cases be of an exceedingly complicated nature, we substitute a multitude of strains and an equal number of stresses, each strain being connected with its corresponding stress by a much more simple relation.

But when all is done, the nature of the connexion between configuration and force remains as mysterious as ever. We can only admit the fact, and if we call all such phenomena phenomena of elasticity, we may find it very convenient to classify them in this way, provided we remember that by the use of the word elasticity we do not profess to explain the cause of the connexion between configuration and energy.

84. Action at a Distance

In the case of the two magnets there is no visible substance connecting the bodies between which the stress exists. The space between the magnets may be filled with air or with water, or we may place the magnets in a vessel and remove the air by an air-pump, till the magnets are left in what is commonly called a vacuum, and yet the mutual action of the magnets will not be altered. We may even place a solid plate of glass or metal or wood between the magnets, and still we find that their mutual action depends simply on their relative position, and is not perceptibly modified by placing any substance between them, unless that substance is one of the magnetic metals. Hence the action between the magnets is commonly spoken of as *action at a distance.*

Attempts have been made, with a certain amount of success[1], to analyse this action at a distance into a continuous distribution of stress in an invisible medium, and thus to establish an analogy between the magnetic action and the action of a spring or a rope in transmitting force; but still the general fact that strains or changes of configuration are accompanied by stresses or internal forces, and that thereby energy is stored up in the system so strained, remains an ultimate fact which has not yet been explained as the result of any more fundamental principle.

85. Theory of Potential Energy more complicated than that of Kinetic Energy

Admitting that the energy of a material system may depend on its configuration, the mode in which it so depends may be much more complicated than the mode in which the kinetic energy depends on the motion of the system. For the kinetic energy may be calculated from the motion of the parts of the system by an invariable method. We multiply the mass of each part by half the square of its velocity, and take the sum of all such products. But the potential energy arising from the mutual action of two parts of the system may depend on the relative position of the parts in a manner which may be different in different instances. Thus when two billiard balls approach each other from a distance, there is no sensible action between them till they come so near one another that certain parts appear to be in contact. To bring the centres of the two balls nearer, the parts in contact must be made to yield, and this requires the expenditure of work.

[1] See Clerk Maxwell's *Treatise on Electricity and Magnetism*, Vol. II, Art. 641. [Modern scrutiny requires a distribution of momentum in the medium, which reveals itself for example in the pressure of radiation, in addition to the stress: cf. appendix to J. H. Poynting's *Collected Papers*. It in fact develops into the guiding tensor principle in the theory of gravitational relativity.]

Hence in this case the potential energy is constant for all distances greater than the distance of first contact, and then rapidly increases when the distance is diminished.

The force between magnets varies with the distance in a very different manner, and in fact we find that it is only by experiment that we can ascertain the form of the relation between the configuration of a system and its potential energy.

86. Application of the Method of Energy to the Calculation of Forces

A complete knowledge of the mode in which the energy of a material system varies when the configuration and motion of the system are made to vary is mathematically equivalent to a knowledge of all the dynamical properties of the system. The mathematical methods by which all the forces and stresses in a moving system are deduced from the single mathematical formula which expresses the energy as a function of the variables have been developed by Lagrange, Hamilton, and other eminent mathematicians, but it would be difficult even to describe them in terms of the elementary ideas to which we restrict ourselves in this book. An outline of these methods is given in my treatise on *Electricity*, Part IV, Chapter V, Article 533*, and the application of these dynamical methods to electromagnetic phenomena is given in the chapters immediately following.

But if we consider only the case of a system at rest it is easy to see how we can ascertain the forces of the system when we know how its energy depends on its configuration.

For let us suppose that an agent external to the system produces a displacement from one configuration to another, then if in the new configuration the system

* Reprinted *infra*, p. 123.

possesses more energy than it did at first, it can have received this increase of energy only from the external agent. This agent must therefore have done an amount of work equal to the increase of energy. It must therefore have exerted force in the direction of the displacement, and the mean value of this force, multiplied into the displacement, must be equal to the work done. Hence the mean value of the force may be found by dividing the increase of energy by the displacement.

If the displacement is large this force may vary considerably during the displacement, so that it may be difficult to calculate its mean value; but since the force depends on the configuration, if we make the displacement smaller and smaller the variation of the force will become smaller and smaller, so that at last the force may be regarded as sensibly constant during the displacement.

If, therefore, we calculate for a given configuration the *rate* at which the energy increases with the displacement, by a method similar to that described in Articles 27, 28, and 33, this rate will be numerically equal to the force exerted by the external agent in the direction of the displacement.

If the energy diminishes instead of increasing as the displacement increases, the system must do work on the external agent, and the force exerted by the external agent must be in the direction opposite to that of displacement.

87. Specification of the [Mode of Action] of Forces

In treatises on dynamics the forces spoken of are usually those exerted by the external agent on the material system. In treatises on electricity, on the other hand, the forces spoken of are usually those exerted by the electrified system against an external agent which prevents the system from moving. It is necessary, therefore, in reading any statement about forces, to ascertain whether the force spoken of is to be regarded from the one point of view or the other.

We may in general avoid any ambiguity by viewing the phenomenon as a whole, and speaking of it as a stress exerted between two points or bodies, and distinguishing it as a tension or a pressure, an attraction or a repulsion, according to its direction. See Article 55.

88. APPLICATION TO A SYSTEM IN MOTION

It thus appears that from a knowledge of the potential energy of a system in every possible configuration we may deduce all the external forces which are required to keep the system in [any given] configuration. If the system is at rest, and if these external forces are the actual forces, the system will remain in equilibrium. If the system is in motion the force acting on each particle is that arising from the connexions of the system (equal and opposite to the external force just calculated), together with any external force which may be applied to it. Hence a complete knowledge of the mode in which the potential energy varies with the configuration would enable us to predict every possible motion of the system under the action of given external forces, provided we were able to overcome the purely mathematical difficulties of the calculation.

89. APPLICATION OF THE METHOD OF ENERGY TO THE INVESTIGATION OF REAL BODIES

When we pass from abstract dynamics to physics—from material systems, whose only properties are those expressed by their definitions, to real bodies, whose properties we have to investigate—we find that there are many phenomena which we are not able to explain as changes in the configuration and motion of a material system.

Of course if we begin by assuming that the real bodies are systems composed of matter which agrees in all respects with the definitions we have laid down, we may go on to assert that all phenomena are changes of configuration and motion, though we are not prepared to define the kind of configuration and motion by which the particular phenomena are to be explained. But in accurate science such asserted explanations must be estimated, not by their promises, but by their performances. The configuration and motion of a system are facts capable of being described in an accurate manner, and therefore, in order that the explanation of a phenomenon by the configuration and motion of a material system may be admitted as an addition to our scientific knowledge, the configurations, motions, and forces must be specified, and shown to be consistent with known facts, as well as capable of accounting for the phenomenon.

90. Variables on which the Energy depends

But even when the phenomena we are studying have not yet been explained dynamically, we are still able to make great use of the principle of the conservation of energy as a guide to our researches.

To apply this principle, we in the first place assume that the quantity of energy in a material system depends on the state of that system, so that for a given state there is a definite amount of energy.

Hence the first step is to define the different states of the system, and when we have to deal with real bodies we must define their state with respect not only to the configuration and motion of their visible parts, but if we have reason to suspect that the configuration and motion of their invisible particles influence the visible phenomenon, we must devise some method of estimating the energy thence arising.

Thus pressure, temperature, electric potential, and chemical composition are variable quantities, the values of which serve to specify the state of a body, and in general the energy of the body depends on the values of these and other variables.

91. Energy in Terms of the Variables

The next step in our investigation is to determine how much work must be done by external agency on the body in order to make it pass from one specified state to another.

For this purpose it is sufficient to know the work required to make the body pass from a particular state, which we may call the *standard state*, into any other specified state. The energy in the latter state is equal to that in the standard state, together with the work required to bring it from the standard state into the specified state. The fact that this work is the same through whatever series of states the system has passed from the standard state to the specified state is the foundation of the whole theory of energy.

Since all the phenomena depend on the variations of the energy of the body, and not on its total value, it is unnecessary, even if it were possible, to form any estimate of the energy of the body in its standard state.

92. Theory of Heat

One of the most important applications of the principle of the conservation of energy is to the investigation of the nature of heat.

At one time it was supposed that the difference between the states of a body when hot and when cold was due to the presence of a substance called caloric, which existed in greater abundance in the body when hot than when cold. But the experiments of Rumford on the heat produced by the friction of metal, and of Davy on the melting of ice by friction, have shown that when work is spent in overcoming friction, the amount of heat produced is proportional to the work spent.

The experiments of Hirn have also shown that when heat is made to do work in a steam-engine, part of the heat disappears, and that the heat which disappears is proportional to the work done.

A very careful measurement of the work spent in friction, and of the heat produced, has been made by Joule, who finds that the heat required to raise one pound of water from 39° F. to 40° F. is equivalent to 772 foot-pounds of work at Manchester, or 24,858 foot-poundals.

From this we may find that the heat required to raise one gramme of water from 3° C. to 4° C. is 42,000,000 ergs.

93. Heat a Form of Energy

Now, since heat can be produced it cannot be a substance; and since whenever mechanical energy is lost by friction there is a production of heat, and whenever there is a gain of mechanical energy in an engine there is a loss of heat; and since the quantity of energy lost or gained is proportional to the quantity of heat gained or lost, we conclude that heat is a form of energy.

We have also reasons for believing that the minute particles of a hot body are in a state of rapid agitation, that is to say, that each particle is always moving very swiftly, but that the direction of its motion alters so often that it makes little or no progress from one region to another.

If this be the case, a part, and it may be a very large part, of the energy of a hot body must be in the form of kinetic energy.

But for our present purpose it is unnecessary to ascertain in what form energy exists in a hot body; the most important fact is that energy may be measured in the form of heat, and since every kind of energy may be converted into heat, this gives us one of the most convenient methods of measuring it.

94. Energy measured as Heat

Thus when certain substances are placed in contact chemical actions take place, the substances combine in a new way, and the new group of substances has different chemical properties from the original group of substances. During this process mechanical work may be done by the expansion of the mixture, as when gunpowder is fired; an electric current may be produced, as in the voltaic battery; and heat may be generated, as in most chemical actions.

The energy given out in the form of mechanical work may be measured directly, or it may be transformed into heat by friction. The energy spent in producing the electric current may be estimated as heat by causing the current to flow through a conductor of such a form that the heat generated in it can easily be measured. Care must be taken that no energy is transmitted to a distance in the form of sound or radiant heat without being duly accounted for.

The energy remaining in the mixture, together with the energy which has escaped, must be equal to the original energy.

Andrews, Favre and Silbermann, [Julius Thomsen,] and others, have measured the quantity of heat produced when a certain quantity of oxygen or of chlorine combines with its equivalent of other substances. These measurements enable us to calculate the excess of the energy which the substances concerned had in their original state, when uncombined, above that which they have after combination.

95. Scientific Work to be done

Though a great deal of excellent work of this kind has already been done, the extent of the field hitherto investigated appears quite insignificant when we consider the boundless variety and complexity of the natural bodies with which we have to deal.

In fact the special work which lies before the physical inquirer in the present state of science is the determination of the quantity of energy which enters or leaves a material system during the passage of the system from its standard state to any other definite state.

96. History of the Doctrine of Energy

The scientific importance of giving a name to the quantity which we call kinetic energy seems to have been first recognised by Leibniz, who gave to the product of the mass by the square of the velocity the name of *Vis Viva*. This is twice the kinetic energy.

Newton, in the "Scholium to the Laws of Motion," expresses the relation between the rate at which work is done by the external agent, and the rate at which it is given out, stored up, or transformed by any machine or other material system, in the following statement, which he makes in order to show the wide extent of the application of the Third Law of Motion.

"If the action of the external agent is estimated by the product of its force into its velocity, and the reaction of the resistance in the same way by the product of the velocity of each part of the system into the resisting force arising from friction, cohesion, weight, and acceleration, the action and reaction will be equal to each other, whatever be the nature and motion of the system." That this statement of Newton's implicitly contains nearly the whole doctrine of energy was first pointed out by Thomson and Tait*.

The words Action and Reaction as they occur in the enunciation of the Third Law of Motion are explained to mean Forces, that is to say, they are the opposite aspects of one and the same Stress.

In the passage quoted above a new and different sense is given to these words by estimating Action and Reaction by the product of a force into the velocity of

* *Treatise on Natural Philosophy*, vol. I, 1867, § 268.

"Newton, in a Scholium to his Third Law of Motion, has stated the relation between work and kinetic energy in a manner so perfect that it cannot be improved, but at the same time with so little apparent effort or desire to attract attention that no one seems to have been struck with the great importance of the passage till it was pointed out recently (1867) by Thomson and Tait." Clerk Maxwell's *Theory of Heat*, ch. IV on "Elementary Dynamical Principles," p. 91.

its point of application. According to this definition the Action of the external agent is the rate at which it does work. This is what is meant by the Power of a steam-engine or other prime mover. It is generally expressed by the estimated number of ideal horses which would be required to do the work at the same rate as the engine, and this is called the Horse-power of the engine.

When we wish to express by a single word the rate at which work is done by an agent we shall call it the Power of the agent, defining the power as the work done in the unit of time.

The use of the term Energy, in a precise and scientific sense, to express the quantity of work which a material system can do, was introduced by Dr Young*.

97. On the Different Forms of Energy

The energy which a body has in virtue of its motion is called kinetic energy.

A system may also have energy in virtue of its configuration, if the forces of the system are such that the system will do work against external resistance while it passes into another configuration. This energy is called Potential Energy. Thus when a stone has been lifted to a certain height above the earth's surface, the system of two bodies, the stone and the earth, has potential energy, and is able to do a certain amount of work during the descent of the stone. This potential energy is due to the fact that the stone and the earth attract each other, so that work has to be spent by the man who lifts the stone and draws it away from the earth, and after the stone is lifted the attraction between the earth and the stone is capable of doing work as the stone descends. This kind of energy, therefore, depends upon the work which the forces of the system would do

* *Lectures on Natural Philosophy* [1807], Lecture VIII.

if the parts of the system were to yield to the action of these forces. This is called the "Sum of the Tensions" by Helmholtz in his celebrated memoir on the "Conservation of Energy."* Thomson called it Statical Energy; it has also been called Energy of Position; but Rankine introduced the term Potential Energy†—a very felicitous expression, since it not only signifies the energy which the system has not in actual possession, but only has the power to acquire, but it also indicates its connexion with what has been called (on other grounds) the Potential Function‡.

The different forms in which energy has been found to exist in material systems have been placed in one or other of these two classes—Kinetic Energy, due to motion, and Potential Energy, due to configuration.

Thus a hot body, by giving out heat to a colder body, may be made to do work by causing the cold body to expand in opposition to pressure. A material system, therefore, in which there is a non-uniform distribution of temperature has the capacity of doing work, or energy. This energy is now believed to be kinetic energy, due to a motion of agitation in the smallest parts of the hot body.

Gunpowder has energy, for when fired it is capable of setting a cannon-ball in motion. The energy of gunpowder is Chemical Energy, arising from the power which the constituents of gunpowder possess of arranging themselves in a new manner when exploded, so as to occupy a much larger volume than the gunpowder does. In the present state of science chemists figure to themselves chemical action as a rearrangement of particles under the action of forces tending to produce

* Berlin, 1847: translated in Taylor's *Scientific Memoirs*, Feb. 1853. [Remarkable mainly for its wide ramifications into electric and chemical theory.]

† The *vis potentialis* of Daniel Bernoulli, as contrasted with *vis viva*, e.g. for the case of a bent spring; cf. Euler, *De Curvis Elasticis*, in Appendix to *Solutio Problematis Isoperimetrici...* (1744).

‡ The term Potential was employed independently by Gauss and by Green, and so probably originated with D. Bernoulli.

this change of arrangement. From this point of view, therefore, chemical energy is potential energy.

Air, compressed in the chamber of an air-gun, is capable of propelling a bullet. The energy of compressed air was at one time supposed to arise from the mutual repulsion of its particles. If this explanation were the true one its energy would be potential energy. In more recent times it has been thought that the particles of the air are in a state of motion, and that its pressure is caused by the impact of these particles on the sides of the vessel. According to this theory the energy of compressed air is kinetic energy.

There are thus many different modes in which a material system may possess energy, and it may be doubtful in some cases whether the energy is of the kinetic or the potential form. The nature of energy, however, is the same in whatever form it may be found. The quantity of energy can always be expressed as equated to that of a body of a definite mass moving with a definite velocity.

Philosophical Magazine
25–29

Taylor & Francis
Taylor & Francis Group

THE

LONDON, EDINBURGH, AND DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE.

[FIFTH SERIES.]

OCTOBER 1897.

XL. *Cathode Rays.*

By J. J. Thomson, *M.A.*, *F.R.S.*, *Cavendish Professor of Experimental Physics, Cambridge**.

THE experiments† discussed in this paper were undertaken in the hope of gaining some information as to the nature of the Cathode Rays. The most diverse opinions are held as to these rays; according to the almost unanimous opinion of German physicists they are due to some process in the æther to which—inasmuch as in a uniform magnetic field their course is circular and not rectilinear—no phenomenon hitherto observed is analogous: another view of these rays is that, so far from being wholly ætherial, they are in fact wholly material, and that they mark the paths of particles of matter charged with negative electricity. It would seem at first sight that it ought not to be difficult to discriminate between views so different, yet experience shows that this is not the case, as amongst the physicists who have most deeply studied the subject can be found supporters of either theory.

The electrified-particle theory has for purposes of research a great advantage over the ætherial theory, since it is definite and its consequences can be predicted; with the ætherial theory it is impossible to predict what will happen under any given circumstances, as on this theory we are dealing with hitherto

* Communicated by the Author.

† Some of these experiments have already been described in a paper read before the Cambridge Philosophical Society (Proceedings, vol. ix. 1897), and in a Friday Evening Discourse at the Royal Institution ('Electrician,' May 21, 1897).

Reprinted from *Philosophical Magazine Series 5*, 44:269, 293–316.

DOI: 10.1080/14786431003659214
http://www.informaworld.com

unobserved phenomena in the æther, of whose laws we are ignorant.

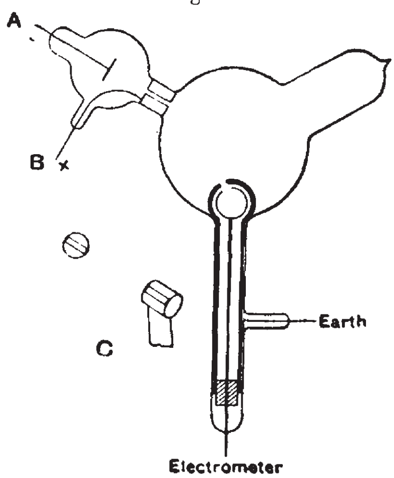
The following experiments were made to test some of the consequences of the electrified-particle theory.

Charge carried by the Cathode Rays.

If these rays are negatively electrified particles, then when they enter an enclosure they ought to carry into it a charge of negative electricity. This has been proved to be the case by Perrin, who placed in front of a plane cathode two coaxial metallic cylinders which were insulated from each other: the outer of these cylinders was connected with the earth, the inner with a gold-leaf electroscope. These cylinders were closed except for two small holes, one in each cylinder, placed so that the cathode rays could pass through them into the inside of the inner cylinder. Perrin found that when the rays passed into the inner cylinder the electroscope received a charge of negative electricity, while no charge went to the electroscope when the rays were deflected by a magnet so as no longer to pass through the hole.

This experiment proves that something charged with negative electricity is shot off from the cathode, travelling at right angles to it, and that this something is deflected by a magnet; it is open, however, to the objection that it does not prove that the cause of the electrification in the electroscope has anything to do with the cathode rays. Now the supporters of the ætherial theory do not deny that electrified particles are shot off from the cathode; they deny, however, that these charged particles have any more to do with the cathode rays than a rifle-ball has with the flash when a rifle is fired. I have therefore repeated Perrin's experiment in a form which is not open to this objection. The arrangement used was as follows:—Two coaxial cylinders (fig. 1) with slits in them are placed in a bulb connected with the discharge-tube; the cathode rays from the cathode A pass into the bulb through a slit in a metal plug fitted into the neck of the tube; this plug is connected with the anode and is put to earth. The cathode rays thus do not fall upon the cylinders unless they are deflected by a magnet. The outer cylinder is connected with the earth, the inner with the electrometer. When the cathode rays (whose path was traced by the phosphorescence on the glass) did not fall on the slit, the electrical charge sent to the electrometer when the induction-coil producing the rays was set in action was small and irregular; when, however, the rays were bent by a magnet so as to fall on the slit there was a large charge of negative electricity sent to the electrometer. I was surprised at the magnitude of the charge; on some occasions enough negative electricity went through the narrow slit into the inner cylinder in one second to alter the potential of a capacity of 1·5 microfarads by 20 volts. If the rays were so

Fig. 1.



much bent by the magnet that they overshot the slits in the cylinder, the charge passing into the cylinder fell again to a very small fraction of its value when the aim was true. Thus this experiment shows that however we twist and deflect the cathode rays by magnetic forces, the negative electrification follows the same path as the rays, and that this negative electrification is indissolubly connected with the cathode rays.

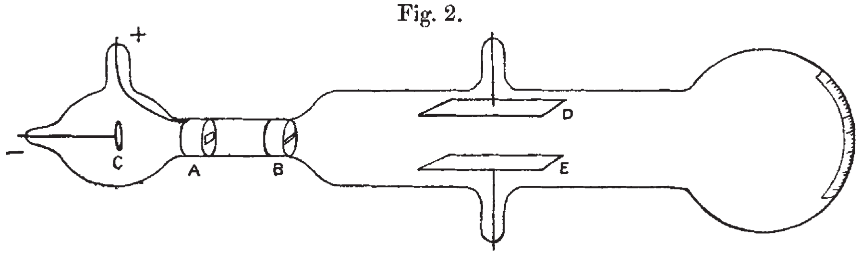
When the rays are turned by the magnet so as to pass through the slit into the inner cylinder, the deflexion of the electrometer connected with this cylinder increases up to a certain value, and then remains stationary although the rays continue to pour into the cylinder. This is due to the fact that the gas in the bulb becomes a conductor of electricity when the cathode rays pass through it, and thus, though the inner cylinder is perfectly insulated when the rays are not passing, yet as soon as the rays pass through the bulb the air between the inner cylinder and the outer one becomes a conductor, and the electricity escapes from the inner cylinder to the earth. Thus the charge within the inner cylinder does not go on continually increasing; the cylinder settles down into a state of equilibrium in which the rate at which it gains negative electricity from the rays is equal to the rate at which it loses it by conduction through the air. If the inner cylinder has initially a positive charge it rapidly loses that charge and acquires a negative one; while if the initial charge is a negative one, the cylinder will leak if the initial negative potential is numerically greater than the equilibrium value.

Deflexion of the Cathode Rays by an Electrostatic Field.

An objection very generally urged against the view that the cathode rays are negatively electrified particles, is that hitherto no deflexion of the rays has been observed under a small electrostatic force, and though the rays are deflected when they pass near electrodes connected with sources of large differences of potential, such as induction-coils or electrical machines, the deflexion in this case is regarded by the supporters of the ætherial theory as due to the discharge passing between the electrodes, and not primarily to the electrostatic field. Hertz made the rays travel between two parallel plates of metal placed inside the discharge-tube, but found that they were not deflected when the plates were connected with a battery of storage-cells; on repeating this experiment I at first got the same result, but subsequent experiments showed that the absence of deflexion is due to the conductivity conferred on the rarefied gas by the cathode rays. On measuring this conductivity it was found that it diminished very rapidly as the exhaustion increased; it seemed then that on trying Hertz's experiment at very high exhaustions there might be a chance of detecting the deflexion of the cathode rays by an electrostatic force.

The apparatus used is represented in fig. 2.

Fig. 2.



The rays from the cathode C pass through a slit in the anode A, which is a metal plug fitting tightly into the tube and connected with the earth; after passing through a second slit in another earth-connected metal plug B, they travel between two parallel aluminium plates about 5 cm. long by 2 broad and at a distance of 1·5 cm. apart; they then fall on the end of the tube and produce a narrow well-defined phosphorescent patch. A scale pasted on the outside of the tube serves to measure the deflexion of this patch.

At high exhaustions the rays were deflected when the two aluminium plates were connected with the terminals of a battery of small storage-cells; the rays were depressed when the upper plate was connected with the negative pole of the battery, the lower with the positive, and raised when the upper plate was connected with the positive, the lower with the negative pole. The deflexion was proportional to the difference of potential between the plates, and I could detect the deflexion when the potential-difference was as small as two volts. It was only when the vacuum was a good one that the deflexion took place, but that the absence of deflexion is due to the conductivity of the medium is shown by what takes place when the vacuum has just arrived at the stage at which the deflexion begins. At this stage there is a deflexion of the rays when the plates are first connected with the terminals of the battery, but if this connexion is maintained the patch of phosphorescence gradually creeps back to its undeflected position. This is just what would happen if the space between the plates were a conductor, though a very bad one, for then the positive and negative ions between the plates would slowly diffuse, until the positive plate became coated with negative ions, the negative plate with positive ones; thus the electric intensity between the plates would vanish and the cathode rays be free from electrostatic force. Another illustration of this is afforded by what happens when the pressure is low enough to show the deflexion and a large difference of potential, say 200 volts, is established between the plates; under these circumstances there is a large deflexion of the cathode rays, but the medium under the large electromotive force breaks down every now and then and a bright discharge passes between the plates; when this occurs the phosphorescent patch produced by the cathode rays jumps back to its undeflected position. When the cathode rays are deflected by the electrostatic field, the phosphorescent band breaks up into several bright bands separated by comparatively dark spaces; the phenomena are exactly analogous to those observed by Birkeland when the cathode rays are deflected by a magnet, and called by him the magnetic spectrum.

A series of measurements of the deflexion of the rays by the electrostatic force under various circumstances will be found later on in the part of the paper which deals with the velocity of the rays and the ratio of the mass of the electrified particles to the charge carried by them. It may, however, be mentioned here that the deflexion gets smaller as the pressure diminishes, and when in consequence the potential-difference in the tube in the neighbourhood of the cathode increases.

ON THE ELECTRODYNAMICS OF MOVING BODIES

By A. EINSTEIN

June 30, 1905

It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case.

Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the "light medium," suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good.[1] We will raise this conjecture (the purport of which will hereafter be called the "Principle of Relativity") to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body. These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies. The introduction of a "luminiferous ether" will prove to be superfluous inasmuch as the view here to be developed will not require an "absolutely stationary space" provided with special properties, nor

[1]The preceding memoir by Lorentz was not at this time known to the author.

assign a velocity-vector to a point of the empty space in which electromagnetic processes take place.

The theory to be developed is based—like all electrodynamics—on the kinematics of the rigid body, since the assertions of any such theory have to do with the relationships between rigid bodies (systems of co-ordinates), clocks, and electromagnetic processes. Insufficient consideration of this circumstance lies at the root of the difficulties which the electrodynamics of moving bodies at present encounters.

I. KINEMATICAL PART

§ 1. Definition of Simultaneity

Let us take a system of co-ordinates in which the equations of Newtonian mechanics hold good.[2] In order to render our presentation more precise and to distinguish this system of co-ordinates verbally from others which will be introduced hereafter, we call it the "stationary system."

If a material point is at rest relatively to this system of co-ordinates, its position can be defined relatively thereto by the employment of rigid standards of measurement and the methods of Euclidean geometry, and can be expressed in Cartesian co-ordinates.

If we wish to describe the *motion* of a material point, we give the values of its co-ordinates as functions of the time. Now we must bear carefully in mind that a mathematical description of this kind has no physical meaning unless we are quite clear as to what we understand by "time." We have to take into account that all our judgments in which time plays a part are always judgments of *simultaneous events*. If, for instance, I say, "That train arrives here at 7 o'clock," I mean something like this: "The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous events."[3]

It might appear possible to overcome all the difficulties attending the definition of "time" by substituting "the position of the small hand of my watch" for "time." And in fact such a definition is satisfactory when we are concerned with defining a time exclusively for the place where the watch is located; but it is no longer satisfactory when we have to connect in time series of events occurring at different places, or—what comes to the same thing—to evaluate the times of events occurring at places remote from the watch.

We might, of course, content ourselves with time values determined by an observer stationed together with the watch at the origin of the co-ordinates, and co-ordinating the corresponding positions of the hands with light signals, given out by every event to be timed, and reaching him through empty space. But this co-ordination has the disadvantage that it is not independent of the standpoint of the observer with the watch or clock, as we know from experience.

[2]i.e. to the first approximation.

[3]We shall not here discuss the inexactitude which lurks in the concept of simultaneity of two events at approximately the same place, which can only be removed by an abstraction.

We arrive at a much more practical determination along the following line of thought.

If at the point A of space there is a clock, an observer at A can determine the time values of events in the immediate proximity of A by finding the positions of the hands which are simultaneous with these events. If there is at the point B of space another clock in all respects resembling the one at A, it is possible for an observer at B to determine the time values of events in the immediate neighbourhood of B. But it is not possible without further assumption to compare, in respect of time, an event at A with an event at B. We have so far defined only an "A time" and a "B time." We have not defined a common "time" for A and B, for the latter cannot be defined at all unless we establish *by definition* that the "time" required by light to travel from A to B equals the "time" it requires to travel from B to A. Let a ray of light start at the "A time" $t_{\rm A}$ from A towards B, let it at the "B time" $t_{\rm B}$ be reflected at B in the direction of A, and arrive again at A at the "A time" $t'_{\rm A}$.

In accordance with definition the two clocks synchronize if

$$t_{\rm B} - t_{\rm A} = t'_{\rm A} - t_{\rm B}.$$

We assume that this definition of synchronism is free from contradictions, and possible for any number of points; and that the following relations are universally valid:—

1. If the clock at B synchronizes with the clock at A, the clock at A synchronizes with the clock at B.

2. If the clock at A synchronizes with the clock at B and also with the clock at C, the clocks at B and C also synchronize with each other.

Thus with the help of certain imaginary physical experiments we have settled what is to be understood by synchronous stationary clocks located at different places, and have evidently obtained a definition of "simultaneous," or "synchronous," and of "time." The "time" of an event is that which is given simultaneously with the event by a stationary clock located at the place of the event, this clock being synchronous, and indeed synchronous for all time determinations, with a specified stationary clock.

In agreement with experience we further assume the quantity

$$\frac{2\rm AB}{t'_A - t_A} = c,$$

to be a universal constant—the velocity of light in empty space.

It is essential to have time defined by means of stationary clocks in the stationary system, and the time now defined being appropriate to the stationary system we call it "the time of the stationary system."

§ 2. On the Relativity of Lengths and Times

The following reflexions are based on the principle of relativity and on the principle of the constancy of the velocity of light. These two principles we define as follows:—

1. The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of co-ordinates in uniform translatory motion.

2. Any ray of light moves in the "stationary" system of co-ordinates with the determined velocity c, whether the ray be emitted by a stationary or by a moving body. Hence

$$\text{velocity} = \frac{\text{light path}}{\text{time interval}}$$

where time interval is to be taken in the sense of the definition in § 1.

Let there be given a stationary rigid rod; and let its length be l as measured by a measuring-rod which is also stationary. We now imagine the axis of the rod lying along the axis of x of the stationary system of co-ordinates, and that a uniform motion of parallel translation with velocity v along the axis of x in the direction of increasing x is then imparted to the rod. We now inquire as to the length of the moving rod, and imagine its length to be ascertained by the following two operations:—

(a) The observer moves together with the given measuring-rod and the rod to be measured, and measures the length of the rod directly by superposing the measuring-rod, in just the same way as if all three were at rest.

(b) By means of stationary clocks set up in the stationary system and synchronizing in accordance with § 1, the observer ascertains at what points of the stationary system the two ends of the rod to be measured are located at a definite time. The distance between these two points, measured by the measuring-rod already employed, which in this case is at rest, is also a length which may be designated "the length of the rod."

In accordance with the principle of relativity the length to be discovered by the operation (a)—we will call it "the length of the rod in the moving system"—must be equal to the length l of the stationary rod.

The length to be discovered by the operation (b) we will call "the length of the (moving) rod in the stationary system." This we shall determine on the basis of our two principles, and we shall find that it differs from l.

Current kinematics tacitly assumes that the lengths determined by these two operations are precisely equal, or in other words, that a moving rigid body at the epoch t may in geometrical respects be perfectly represented by *the same* body *at rest* in a definite position.

We imagine further that at the two ends A and B of the rod, clocks are placed which synchronize with the clocks of the stationary system, that is to say that their indications correspond at any instant to the "time of the stationary system" at the places where they happen to be. These clocks are therefore "synchronous in the stationary system."

We imagine further that with each clock there is a moving observer, and that these observers apply to both clocks the criterion established in § 1 for the synchronization of two clocks. Let a ray of light depart from A at the time[4] $t_{\rm A}$,

[4] "Time" here denotes "time of the stationary system" and also "position of hands of the moving clock situated at the place under discussion."

let it be reflected at B at the time t_B, and reach A again at the time t'_A. Taking into consideration the principle of the constancy of the velocity of light we find that

$$t_\mathrm{B} - t_\mathrm{A} = \frac{r_\mathrm{AB}}{c - v} \text{ and } t'_\mathrm{A} - t_\mathrm{B} = \frac{r_\mathrm{AB}}{c + v}$$

where r_AB denotes the length of the moving rod—measured in the stationary system. Observers moving with the moving rod would thus find that the two clocks were not synchronous, while observers in the stationary system would declare the clocks to be synchronous.

So we see that we cannot attach any *absolute* signification to the concept of simultaneity, but that two events which, viewed from a system of co-ordinates, are simultaneous, can no longer be looked upon as simultaneous events when envisaged from a system which is in motion relatively to that system.

§ 3. Theory of the Transformation of Co-ordinates and Times from a Stationary System to another System in Uniform Motion of Translation Relatively to the Former

Let us in "stationary" space take two systems of co-ordinates, i.e. two systems, each of three rigid material lines, perpendicular to one another, and issuing from a point. Let the axes of X of the two systems coincide, and their axes of Y and Z respectively be parallel. Let each system be provided with a rigid measuring-rod and a number of clocks, and let the two measuring-rods, and likewise all the clocks of the two systems, be in all respects alike.

Now to the origin of one of the two systems (k) let a constant velocity v be imparted in the direction of the increasing x of the other stationary system (K), and let this velocity be communicated to the axes of the co-ordinates, the relevant measuring-rod, and the clocks. To any time of the stationary system K there then will correspond a definite position of the axes of the moving system, and from reasons of symmetry we are entitled to assume that the motion of k may be such that the axes of the moving system are at the time t (this "t" always denotes a time of the stationary system) parallel to the axes of the stationary system.

We now imagine space to be measured from the stationary system K by means of the stationary measuring-rod, and also from the moving system k by means of the measuring-rod moving with it; and that we thus obtain the co-ordinates x, y, z, and ξ, η, ζ respectively. Further, let the time t of the stationary system be determined for all points thereof at which there are clocks by means of light signals in the manner indicated in § 1; similarly let the time τ of the moving system be determined for all points of the moving system at which there are clocks at rest relatively to that system by applying the method, given in § 1, of light signals between the points at which the latter clocks are located.

To any system of values x, y, z, t, which completely defines the place and time of an event in the stationary system, there belongs a system of values ξ,

η, ζ τ, determining that event relatively to the system k, and our task is now to find the system of equations connecting these quantities.

In the first place it is clear that the equations must be *linear* on account of the properties of homogeneity which we attribute to space and time.

If we place $x' = x - vt$, it is clear that a point at rest in the system k must have a system of values x', y, z, independent of time. We first define τ as a function of x', y, z, and t. To do this we have to express in equations that τ is nothing else than the summary of the data of clocks at rest in system k, which have been synchronized according to the rule given in § 1.

From the origin of system k let a ray be emitted at the time τ_0 along the X-axis to x', and at the time τ_1 be reflected thence to the origin of the co-ordinates, arriving there at the time τ_2; we then must have $\frac{1}{2}(\tau_0 + \tau_2) = \tau_1$, or, by inserting the arguments of the function τ and applying the principle of the constancy of the velocity of light in the stationary system:—

$$\frac{1}{2}\left[\tau(0,0,0,t) + \tau\left(0,0,0,t + \frac{x'}{c-v} + \frac{x'}{c+v}\right)\right] = \tau\left(x',0,0,t + \frac{x'}{c-v}\right).$$

Hence, if x' be chosen infinitesimally small,

$$\frac{1}{2}\left(\frac{1}{c-v} + \frac{1}{c+v}\right)\frac{\partial\tau}{\partial t} = \frac{\partial\tau}{\partial x'} + \frac{1}{c-v}\frac{\partial\tau}{\partial t},$$

or

$$\frac{\partial\tau}{\partial x'} + \frac{v}{c^2 - v^2}\frac{\partial\tau}{\partial t} = 0.$$

It is to be noted that instead of the origin of the co-ordinates we might have chosen any other point for the point of origin of the ray, and the equation just obtained is therefore valid for all values of x', y, z.

An analogous consideration—applied to the axes of Y and Z—it being borne in mind that light is always propagated along these axes, when viewed from the stationary system, with the velocity $\sqrt{c^2 - v^2}$ gives us

$$\frac{\partial\tau}{\partial y} = 0, \frac{\partial\tau}{\partial z} = 0.$$

Since τ is a *linear* function, it follows from these equations that

$$\tau = a\left(t - \frac{v}{c^2 - v^2}x'\right)$$

where a is a function $\phi(v)$ at present unknown, and where for brevity it is assumed that at the origin of k, $\tau = 0$, when $t = 0$.

With the help of this result we easily determine the quantities ξ, η, ζ by expressing in equations that light (as required by the principle of the constancy of the velocity of light, in combination with the principle of relativity) is also propagated with velocity c when measured in the moving system. For a ray of light emitted at the time $\tau = 0$ in the direction of the increasing ξ

$$\xi = c\tau \text{ or } \xi = ac\left(t - \frac{v}{c^2 - v^2}x'\right).$$

But the ray moves relatively to the initial point of k, when measured in the stationary system, with the velocity $c - v$, so that

$$\frac{x'}{c - v} = t.$$

If we insert this value of t in the equation for ξ, we obtain

$$\xi = a\frac{c^2}{c^2 - v^2}x'.$$

In an analogous manner we find, by considering rays moving along the two other axes, that

$$\eta = c\tau = ac\left(t - \frac{v}{c^2 - v^2}x'\right)$$

when

$$\frac{y}{\sqrt{c^2 - v^2}} = t,\ x' = 0.$$

Thus

$$\eta = a\frac{c}{\sqrt{c^2 - v^2}}y \text{ and } \zeta = a\frac{c}{\sqrt{c^2 - v^2}}z.$$

Substituting for x' its value, we obtain

$$\begin{aligned}
\tau &= \phi(v)\beta(t - vx/c^2),\\
\xi &= \phi(v)\beta(t - vt),\\
\eta &= \phi(v)y,\\
\zeta &= \phi(v)z,
\end{aligned}$$

where

$$\beta = \frac{1}{\sqrt{1 - v^2/c^2}},$$

and ϕ is an as yet unknown function of v. If no assumption whatever be made as to the initial position of the moving system and as to the zero point of τ, an additive constant is to be placed on the right side of each of these equations.

We now have to prove that any ray of light, measured in the moving system, is propagated with the velocity c, if, as we have assumed, this is the case in the stationary system; for we have not as yet furnished the proof that the principle of the constancy of the velocity of light is compatible with the principle of relativity.

At the time $t = \tau = 0$, when the origin of the co-ordinates is common to the two systems, let a spherical wave be emitted therefrom, and be propagated with the velocity c in system K. If (x, y, z) be a point just attained by this wave, then

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

Transforming this equation with the aid of our equations of transformation we obtain after a simple calculation

$$\xi^2 + \eta^2 + \zeta^2 = c^2 \tau^2.$$

The wave under consideration is therefore no less a spherical wave with velocity of propagation c when viewed in the moving system. This shows that our two fundamental principles are compatible.[5]

In the equations of transformation which have been developed there enters an unknown function ϕ of v, which we will now determine.

For this purpose we introduce a third system of co-ordinates K′, which relatively to the system k is in a state of parallel translatory motion parallel to the axis of X, such that the origin of co-ordinates of system k moves with velocity $-v$ on the axis of X. At the time $t = 0$ let all three origins coincide, and when $t = x = y = z = 0$ let the time t' of the system K′ be zero. We call the co-ordinates, measured in the system K′, x', y', z', and by a twofold application of our equations of transformation we obtain

$$\begin{array}{lllll} t' & = & \phi(-v)\beta(-v)(\tau + v\xi/c^2) & = & \phi(v)\phi(-v)t, \\ x' & = & \phi(-v)\beta(-v)(\xi + v\tau) & = & \phi(v)\phi(-v)x, \\ y' & = & \phi(-v)\eta & = & \phi(v)\phi(-v)y, \\ z' & = & \phi(-v)\zeta & = & \phi(v)\phi(-v)z. \end{array}$$

Since the relations between x', y', z' and x, y, z do not contain the time t, the systems K and K′ are at rest with respect to one another, and it is clear that the transformation from K to K′ must be the identical transformation. Thus

$$\phi(v)\phi(-v) = 1.$$

We now inquire into the signification of $\phi(v)$. We give our attention to that part of the axis of Y of system k which lies between $\xi = 0, \eta = 0, \zeta = 0$ and $\xi = 0, \eta = l, \zeta = 0$. This part of the axis of Y is a rod moving perpendicularly

[5]The equations of the Lorentz transformation may be more simply deduced directly from the condition that in virtue of those equations the relation $x^2 + y^2 + z^2 = c^2t^2$ shall have as its consequence the second relation $\xi^2 + \eta^2 + \zeta^2 = c^2\tau^2$.

to its axis with velocity v relatively to system K. Its ends possess in K the co-ordinates

$$x_1 = vt,\ y_1 = \frac{l}{\phi(v)},\ z_1 = 0$$

and

$$x_2 = vt,\ y_2 = 0,\ z_2 = 0.$$

The length of the rod measured in K is therefore $l/\phi(v)$; and this gives us the meaning of the function $\phi(v)$. From reasons of symmetry it is now evident that the length of a given rod moving perpendicularly to its axis, measured in the stationary system, must depend only on the velocity and not on the direction and the sense of the motion. The length of the moving rod measured in the stationary system does not change, therefore, if v and $-v$ are interchanged. Hence follows that $l/\phi(v) = l/\phi(-v)$, or

$$\phi(v) = \phi(-v).$$

It follows from this relation and the one previously found that $\phi(v) = 1$, so that the transformation equations which have been found become

$$\begin{aligned} \tau &= \beta(t - vx/c^2), \\ \xi &= \beta(x - vt), \\ \eta &= y, \\ \zeta &= z, \end{aligned}$$

where

$$\beta = 1/\sqrt{1 - v^2/c^2}.$$

§ 4. Physical Meaning of the Equations Obtained in Respect to Moving Rigid Bodies and Moving Clocks

We envisage a rigid sphere[6] of radius R, at rest relatively to the moving system k, and with its centre at the origin of co-ordinates of k. The equation of the surface of this sphere moving relatively to the system K with velocity v is

$$\xi^2 + \eta^2 + \zeta^2 = \mathrm{R}^2.$$

The equation of this surface expressed in x, y, z at the time $t = 0$ is

$$\frac{x^2}{(\sqrt{1 - v^2/c^2})^2} + y^2 + z^2 = \mathrm{R}^2.$$

[6]That is, a body possessing spherical form when examined at rest.

A rigid body which, measured in a state of rest, has the form of a sphere, therefore has in a state of motion—viewed from the stationary system—the form of an ellipsoid of revolution with the axes

$$\mathrm{R}\sqrt{1-v^2/c^2},\ \mathrm{R},\ \mathrm{R}.$$

Thus, whereas the Y and Z dimensions of the sphere (and therefore of every rigid body of no matter what form) do not appear modified by the motion, the X dimension appears shortened in the ratio $1:\sqrt{1-v^2/c^2}$, i.e. the greater the value of v, the greater the shortening. For $v=c$ all moving objects—viewed from the "stationary" system—shrivel up into plane figures.† For velocities greater than that of light our deliberations become meaningless; we shall, however, find in what follows, that the velocity of light in our theory plays the part, physically, of an infinitely great velocity.

It is clear that the same results hold good of bodies at rest in the "stationary" system, viewed from a system in uniform motion.

Further, we imagine one of the clocks which are qualified to mark the time t when at rest relatively to the stationary system, and the time τ when at rest relatively to the moving system, to be located at the origin of the co-ordinates of k, and so adjusted that it marks the time τ. What is the rate of this clock, when viewed from the stationary system?

Between the quantities x, t, and τ, which refer to the position of the clock, we have, evidently, $x=vt$ and

$$\tau=\frac{1}{\sqrt{1-v^2/c^2}}(t-vx/c^2).$$

Therefore,

$$\tau=t\sqrt{1-v^2/c^2}=t-(1-\sqrt{1-v^2/c^2})t$$

whence it follows that the time marked by the clock (viewed in the stationary system) is slow by $1-\sqrt{1-v^2/c^2}$ seconds per second, or—neglecting magnitudes of fourth and higher order—by $\frac{1}{2}v^2/c^2$.

From this there ensues the following peculiar consequence. If at the points A and B of K there are stationary clocks which, viewed in the stationary system, are synchronous; and if the clock at A is moved with the velocity v along the line AB to B, then on its arrival at B the two clocks no longer synchronize, but the clock moved from A to B lags behind the other which has remained at B by $\frac{1}{2}tv^2/c^2$ (up to magnitudes of fourth and higher order), t being the time occupied in the journey from A to B.

It is at once apparent that this result still holds good if the clock moves from A to B in any polygonal line, and also when the points A and B coincide.

If we assume that the result proved for a polygonal line is also valid for a continuously curved line, we arrive at this result: If one of two synchronous

†Editor's note: In the original 1923 English edition, this phrase was erroneously translated as "plain figures". I have used the correct "plane figures" in this edition.

clocks at A is moved in a closed curve with constant velocity until it returns to A, the journey lasting t seconds, then by the clock which has remained at rest the travelled clock on its arrival at A will be $\frac{1}{2}tv^2/c^2$ second slow. Thence we conclude that a balance-clock[7] at the equator must go more slowly, by a very small amount, than a precisely similar clock situated at one of the poles under otherwise identical conditions.

§ 5. The Composition of Velocities

In the system k moving along the axis of X of the system K with velocity v, let a point move in accordance with the equations

$$\xi = w_\xi \tau, \eta = w_\eta \tau, \zeta = 0,$$

where w_ξ and w_η denote constants.

Required: the motion of the point relatively to the system K. If with the help of the equations of transformation developed in § 3 we introduce the quantities x, y, z, t into the equations of motion of the point, we obtain

$$\begin{aligned} x &= \frac{w_\xi + v}{1 + v w_\xi / c^2} t, \\ y &= \frac{\sqrt{1 - v^2/c^2}}{1 + v w_\xi / c^2} w_\eta t, \\ z &= 0. \end{aligned}$$

Thus the law of the parallelogram of velocities is valid according to our theory only to a first approximation. We set

$$\begin{aligned} V^2 &= \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2, \\ w^2 &= w_\xi^2 + w_\eta^2, \\ a &= \tan^{-1} w_y / w_x, \end{aligned}$$

a is then to be looked upon as the angle between the velocities v and w. After a simple calculation we obtain

$$V = \frac{\sqrt{(v^2 + w^2 + 2vw\cos a) - (vw\sin a/c^2)^2}}{1 + vw\cos a/c^2}.$$

[7] Not a pendulum-clock, which is physically a system to which the Earth belongs. This case had to be excluded.

It is worthy of remark that v and w enter into the expression for the resultant velocity in a symmetrical manner. If w also has the direction of the axis of X, we get

$$V = \frac{v+w}{1+vw/c^2}.$$

It follows from this equation that from a composition of two velocities which are less than c, there always results a velocity less than c. For if we set $v = c-\kappa, w = c-\lambda$, κ and λ being positive and less than c, then

$$V = c\frac{2c-\kappa-\lambda}{2c-\kappa-\lambda+\kappa\lambda/c} < c.$$

It follows, further, that the velocity of light c cannot be altered by composition with a velocity less than that of light. For this case we obtain

$$V = \frac{c+w}{1+w/c} = c.$$

We might also have obtained the formula for V, for the case when v and w have the same direction, by compounding two transformations in accordance with § 3. If in addition to the systems K and k figuring in § 3 we introduce still another system of co-ordinates k' moving parallel to k, its initial point moving on the axis of X with the velocity w, we obtain equations between the quantities x, y, z, t and the corresponding quantities of k', which differ from the equations found in § 3 only in that the place of "v" is taken by the quantity

$$\frac{v+w}{1+vw/c^2};$$

from which we see that such parallel transformations—necessarily—form a group.

We have now deduced the requisite laws of the theory of kinematics corresponding to our two principles, and we proceed to show their application to electrodynamics.

II. ELECTRODYNAMICAL PART

§ 6. Transformation of the Maxwell-Hertz Equations for Empty Space. On the Nature of the Electromotive Forces Occurring in a Magnetic Field During Motion

Let the Maxwell-Hertz equations for empty space hold good for the stationary system K, so that we have

$$\begin{aligned}
\frac{1}{c}\frac{\partial \mathrm{X}}{\partial t} &= \frac{\partial \mathrm{N}}{\partial y} - \frac{\partial \mathrm{M}}{\partial z}, & \frac{1}{c}\frac{\partial \mathrm{L}}{\partial t} &= \frac{\partial \mathrm{Y}}{\partial z} - \frac{\partial \mathrm{Z}}{\partial y}, \\
\frac{1}{c}\frac{\partial \mathrm{Y}}{\partial t} &= \frac{\partial \mathrm{L}}{\partial z} - \frac{\partial \mathrm{N}}{\partial x}, & \frac{1}{c}\frac{\partial \mathrm{M}}{\partial t} &= \frac{\partial \mathrm{Z}}{\partial x} - \frac{\partial \mathrm{X}}{\partial z}, \\
\frac{1}{c}\frac{\partial \mathrm{Z}}{\partial t} &= \frac{\partial \mathrm{M}}{\partial x} - \frac{\partial \mathrm{L}}{\partial y}, & \frac{1}{c}\frac{\partial \mathrm{N}}{\partial t} &= \frac{\partial \mathrm{X}}{\partial y} - \frac{\partial \mathrm{Y}}{\partial x},
\end{aligned}$$

where (X, Y, Z) denotes the vector of the electric force, and (L, M, N) that of the magnetic force.

If we apply to these equations the transformation developed in § 3, by referring the electromagnetic processes to the system of co-ordinates there introduced, moving with the velocity v, we obtain the equations

$$\begin{aligned}
\tfrac{1}{c}\tfrac{\partial \mathrm{X}}{\partial \tau} &= \tfrac{\partial}{\partial \eta}\left\{\beta\left(\mathrm{N}-\tfrac{v}{c}\mathrm{Y}\right)\right\} &&-\tfrac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{M}+\tfrac{v}{c}\mathrm{Z}\right)\right\},\\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{Y}-\tfrac{v}{c}\mathrm{N}\right)\right\} &= \tfrac{\partial \mathrm{L}}{\partial \xi} &&-\tfrac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{N}-\tfrac{v}{c}\mathrm{Y}\right)\right\},\\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{Z}-\tfrac{v}{c}\mathrm{M}\right)\right\} &= \tfrac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{M}-\tfrac{v}{c}\mathrm{Z}\right)\right\} &&-\tfrac{\partial \mathrm{L}}{\partial \eta},\\
\tfrac{1}{c}\tfrac{\partial \mathrm{L}}{\partial \tau} &= \tfrac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{Y}-\tfrac{v}{c}\mathrm{N}\right)\right\} &&-\tfrac{\partial}{\partial \eta}\left\{\beta\left(\mathrm{Z}-\tfrac{v}{c}\mathrm{M}\right)\right\},\\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{M}-\tfrac{v}{c}\mathrm{Z}\right)\right\} &= \tfrac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{Z}-\tfrac{v}{c}\mathrm{M}\right)\right\} &&-\tfrac{\partial \mathrm{X}}{\partial \zeta},\\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{N}-\tfrac{v}{c}\mathrm{Y}\right)\right\} &= \tfrac{\partial \mathrm{X}}{\partial \eta} &&-\tfrac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{Y}-\tfrac{v}{c}\mathrm{N}\right)\right\},
\end{aligned}$$

where

$$\beta = 1/\sqrt{1-v^2/c^2}.$$

Now the principle of relativity requires that if the Maxwell-Hertz equations for empty space hold good in system K, they also hold good in system k; that is to say that the vectors of the electric and the magnetic force—(X′, Y′, Z′) and (L′, M′, N′)—of the moving system k, which are defined by their ponderomotive effects on electric or magnetic masses respectively, satisfy the following equations:—

$$\begin{aligned}
\frac{1}{c}\frac{\partial \mathrm{X}'}{\partial \tau} &= \frac{\partial \mathrm{N}'}{\partial \eta} - \frac{\partial \mathrm{M}'}{\partial \zeta}, & \frac{1}{c}\frac{\partial \mathrm{L}'}{\partial \tau} &= \frac{\partial \mathrm{Y}'}{\partial \zeta} - \frac{\partial \mathrm{Z}'}{\partial \eta},\\
\frac{1}{c}\frac{\partial \mathrm{Y}'}{\partial \tau} &= \frac{\partial \mathrm{L}'}{\partial \zeta} - \frac{\partial \mathrm{N}'}{\partial \xi}, & \frac{1}{c}\frac{\partial \mathrm{M}'}{\partial \tau} &= \frac{\partial \mathrm{Z}'}{\partial \xi} - \frac{\partial \mathrm{X}'}{\partial \zeta},\\
\frac{1}{c}\frac{\partial \mathrm{Z}'}{\partial \tau} &= \frac{\partial \mathrm{M}'}{\partial \xi} - \frac{\partial \mathrm{L}'}{\partial \eta}, & \frac{1}{c}\frac{\partial \mathrm{N}'}{\partial \tau} &= \frac{\partial \mathrm{X}'}{\partial \eta} - \frac{\partial \mathrm{Y}'}{\partial \xi}.
\end{aligned}$$

Evidently the two systems of equations found for system k must express exactly the same thing, since both systems of equations are equivalent to the Maxwell-Hertz equations for system K. Since, further, the equations of the two systems agree, with the exception of the symbols for the vectors, it follows that the functions occurring in the systems of equations at corresponding places must agree, with the exception of a factor $\psi(v)$, which is common for all functions of the one system of equations, and is independent of ξ, η, ζ and τ but depends upon v. Thus we have the relations

$$
\begin{aligned}
X' &= \psi(v)X, & L' &= \psi(v)L, \\
Y' &= \psi(v)\beta\left(Y - \frac{v}{c}N\right), & M' &= \psi(v)\beta\left(M - \frac{v}{c}Z\right), \\
Z' &= \psi(v)\beta\left(Z - \frac{v}{c}M\right), & N' &= \psi(v)\beta\left(N - \frac{v}{c}Y\right).
\end{aligned}
$$

If we now form the reciprocal of this system of equations, firstly by solving the equations just obtained, and secondly by applying the equations to the inverse transformation (from k to K), which is characterized by the velocity $-v$, it follows, when we consider that the two systems of equations thus obtained must be identical, that $\psi(v)\psi(-v) = 1$. Further, from reasons of symmetry[8] and therefore

$$\psi(v) = 1,$$

and our equations assume the form

$$
\begin{aligned}
X' &= X, & L' &= L, \\
Y' &= \beta\left(Y - \frac{v}{c}N\right), & M' &= \beta\left(M - \frac{v}{c}Z\right), \\
Z' &= \beta\left(Z - \frac{v}{c}M\right), & N' &= \beta\left(N - \frac{v}{c}Y\right).
\end{aligned}
$$

As to the interpretation of these equations we make the following remarks: Let a point charge of electricity have the magnitude "one" when measured in the stationary system K, i.e. let it when at rest in the stationary system exert a force of one dyne upon an equal quantity of electricity at a distance of one cm. By the principle of relativity this electric charge is also of the magnitude "one" when measured in the moving system. If this quantity of electricity is at rest relatively to the stationary system, then by definition the vector (X, Y, Z) is equal to the force acting upon it. If the quantity of electricity is at rest relatively to the moving system (at least at the relevant instant), then the force acting upon it, measured in the moving system, is equal to the vector (X′, Y′, Z′). Consequently the first three equations above allow themselves to be clothed in words in the two following ways:—

1. If a unit electric point charge is in motion in an electromagnetic field, there acts upon it, in addition to the electric force, an "electromotive force" which, if we neglect the terms multiplied by the second and higher powers of v/c, is equal to the vector-product of the velocity of the charge and the magnetic force, divided by the velocity of light. (Old manner of expression.)

2. If a unit electric point charge is in motion in an electromagnetic field, the force acting upon it is equal to the electric force which is present at the locality of the charge, and which we ascertain by transformation of the field to a system of co-ordinates at rest relatively to the electrical charge. (New manner of expression.)

The analogy holds with "magnetomotive forces." We see that electromotive force plays in the developed theory merely the part of an auxiliary concept,

[8] If, for example, X=Y=Z=L=M=0, and N $\neq 0$, then from reasons of symmetry it is clear that when v changes sign without changing its numerical value, Y′ must also change sign without changing its numerical value.

which owes its introduction to the circumstance that electric and magnetic forces do not exist independently of the state of motion of the system of co-ordinates.

Furthermore it is clear that the asymmetry mentioned in the introduction as arising when we consider the currents produced by the relative motion of a magnet and a conductor, now disappears. Moreover, questions as to the "seat" of electrodynamic electromotive forces (unipolar machines) now have no point.

§ 7. Theory of Doppler's Principle and of Aberration

In the system K, very far from the origin of co-ordinates, let there be a source of electrodynamic waves, which in a part of space containing the origin of co-ordinates may be represented to a sufficient degree of approximation by the equations

$$\begin{array}{ll} \mathrm{X} = \mathrm{X}_0 \sin\Phi, & \mathrm{L} = \mathrm{L}_0 \sin\Phi, \\ \mathrm{Y} = \mathrm{Y}_0 \sin\Phi, & \mathrm{M} = \mathrm{M}_0 \sin\Phi, \\ \mathrm{Z} = \mathrm{Z}_0 \sin\Phi, & \mathrm{N} = \mathrm{N}_0 \sin\Phi, \end{array}$$

where

$$\Phi = \omega\left\{t - \frac{1}{c}(lx + my + nz)\right\}.$$

Here $(\mathrm{X}_0, \mathrm{Y}_0, \mathrm{Z}_0)$ and $(\mathrm{L}_0, \mathrm{M}_0, \mathrm{N}_0)$ are the vectors defining the amplitude of the wave-train, and l, m, n the direction-cosines of the wave-normals. We wish to know the constitution of these waves, when they are examined by an observer at rest in the moving system k.

Applying the equations of transformation found in § 6 for electric and magnetic forces, and those found in § 3 for the co-ordinates and the time, we obtain directly

$$\begin{array}{ll} \mathrm{X}' = \mathrm{X}_0 \sin\Phi', & \mathrm{L}' = \mathrm{L}_0 \sin\Phi', \\ \mathrm{Y}' = \beta(\mathrm{Y}_0 - v\mathrm{N}_0/c)\sin\Phi', & \mathrm{M}' = \beta(\mathrm{M}_0 + v\mathrm{Z}_0/c)\sin\Phi', \\ \mathrm{Z}' = \beta(\mathrm{Z}_0 + v\mathrm{M}_0/c)\sin\Phi', & \mathrm{N}' = \beta(\mathrm{N}_0 - v\mathrm{Y}_0/c)\sin\Phi', \end{array}$$
$$\Phi' = \omega'\left\{\tau - \tfrac{1}{c}(l'\xi + m'\eta + n'\zeta)\right\}$$

where

$$\begin{aligned} \omega' &= \omega\beta(1 - lv/c), \\ l' &= \frac{l - v/c}{1 - lv/c}, \\ m' &= \frac{m}{\beta(1 - lv/c)}, \\ n' &= \frac{n}{\beta(1 - lv/c)}. \end{aligned}$$

From the equation for ω' it follows that if an observer is moving with velocity v relatively to an infinitely distant source of light of frequency ν, in such a way that the connecting line "source-observer" makes the angle ϕ with the velocity of the observer referred to a system of co-ordinates which is at rest relatively to the source of light, the frequency ν' of the light perceived by the observer is given by the equation

$$\nu' = \nu \frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}}.$$

This is Doppler's principle for any velocities whatever. When $\phi = 0$ the equation assumes the perspicuous form

$$\nu' = \nu \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

We see that, in contrast with the customary view, when $v = -c, \nu' = \infty$.

If we call the angle between the wave-normal (direction of the ray) in the moving system and the connecting line "source-observer" ϕ', the equation for l' assumes the form

$$\cos\phi' = \frac{\cos\phi - v/c}{1 - \cos\phi \cdot v/c}.$$

This equation expresses the law of aberration in its most general form. If $\phi = \frac{1}{2}\pi$, the equation becomes simply

$$\cos\phi' = -v/c.$$

We still have to find the amplitude of the waves, as it appears in the moving system. If we call the amplitude of the electric or magnetic force A or A$'$ respectively, accordingly as it is measured in the stationary system or in the moving system, we obtain

$$\mathrm{A}'^2 = \mathrm{A}^2 \frac{(1 - \cos\phi \cdot v/c)^2}{1 - v^2/c^2}$$

which equation, if $\phi = 0$, simplifies into

$$\mathrm{A}'^2 = \mathrm{A}^2 \frac{1 - v/c}{1 + v/c}.$$

It follows from these results that to an observer approaching a source of light with the velocity c, this source of light must appear of infinite intensity.

§ 8. Transformation of the Energy of Light Rays. Theory of the Pressure of Radiation Exerted on Perfect Reflectors

Since $\mathrm{A}^2/8\pi$ equals the energy of light per unit of volume, we have to regard $\mathrm{A}'^2/8\pi$, by the principle of relativity, as the energy of light in the moving system.

Thus ${A'}^2/A^2$ would be the ratio of the "measured in motion" to the "measured at rest" energy of a given light complex, if the volume of a light complex were the same, whether measured in K or in k. But this is not the case. If l, m, n are the direction-cosines of the wave-normals of the light in the stationary system, no energy passes through the surface elements of a spherical surface moving with the velocity of light:—

$$(x - lct)^2 + (y - mct)^2 + (z - nct)^2 = \mathrm{R}^2.$$

We may therefore say that this surface permanently encloses the same light complex. We inquire as to the quantity of energy enclosed by this surface, viewed in system k, that is, as to the energy of the light complex relatively to the system k.

The spherical surface—viewed in the moving system—is an ellipsoidal surface, the equation for which, at the time $\tau = 0$, is

$$(\beta\xi - l\beta\xi v/c)^2 + (\eta - m\beta\xi v/c)^2 + (\zeta - n\beta\xi v/c)^2 = \mathrm{R}^2.$$

If S is the volume of the sphere, and S′ that of this ellipsoid, then by a simple calculation

$$\frac{\mathrm{S}'}{\mathrm{S}} = \frac{\sqrt{1 - v^2/c^2}}{1 - \cos\phi \cdot v/c}.$$

Thus, if we call the light energy enclosed by this surface E when it is measured in the stationary system, and E′ when measured in the moving system, we obtain

$$\frac{\mathrm{E}'}{\mathrm{E}} = \frac{{\mathrm{A}'}^2\mathrm{S}'}{\mathrm{A}^2\mathrm{S}} = \frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}},$$

and this formula, when $\phi = 0$, simplifies into

$$\frac{\mathrm{E}'}{\mathrm{E}} = \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

It is remarkable that the energy and the frequency of a light complex vary with the state of motion of the observer in accordance with the same law.

Now let the co-ordinate plane $\xi = 0$ be a perfectly reflecting surface, at which the plane waves considered in § 7 are reflected. We seek for the pressure of light exerted on the reflecting surface, and for the direction, frequency, and intensity of the light after reflexion.

Let the incidental light be defined by the quantities A, $\cos\phi$, ν (referred to system K). Viewed from k the corresponding quantities are

$$\mathrm{A}' = \mathrm{A}\frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}},$$

$$\cos\phi' = \frac{\cos\phi - v/c}{1-\cos\phi\cdot v/c},$$
$$\nu' = \nu\frac{1-\cos\phi\cdot v/c}{\sqrt{1-v^2/c^2}}.$$

For the reflected light, referring the process to system k, we obtain

$$\mathrm{A}'' = \mathrm{A}'$$
$$\cos\phi'' = -\cos\phi'$$
$$\nu'' = \nu'$$

Finally, by transforming back to the stationary system K, we obtain for the reflected light

$$\mathrm{A}''' = \mathrm{A}''\frac{1+cos\phi''\cdot v/c}{\sqrt{1-v^2/c^2}} = \mathrm{A}\frac{1-2\cos\phi\cdot v/c+v^2/c^2}{1-v^2/c^2},$$
$$\cos\phi''' = \frac{\cos\phi''+v/c}{1+\cos\phi''\cdot v/c} = -\frac{(1+v^2/c^2)\cos\phi-2v/c}{1-2\cos\phi\cdot v/c+v^2/c^2},$$
$$\nu''' = \nu''\frac{1+\cos\phi''\cdot v/c}{\sqrt{1-v^2/c^2}} = \nu\frac{1-2\cos\phi\cdot v/c+v^2/c^2}{1-v^2/c^2}.$$

The energy (measured in the stationary system) which is incident upon unit area of the mirror in unit time is evidently $\mathrm{A}^2(c\cos\phi - v)/8\pi$. The energy leaving the unit of surface of the mirror in the unit of time is $\mathrm{A}'''^2(-c\cos\phi''' + v)/8\pi$. The difference of these two expressions is, by the principle of energy, the work done by the pressure of light in the unit of time. If we set down this work as equal to the product Pv, where P is the pressure of light, we obtain

$$\mathrm{P} = 2\cdot\frac{\mathrm{A}^2}{8\pi}\frac{(\cos\phi - v/c)^2}{1-v^2/c^2}.$$

In agreement with experiment and with other theories, we obtain to a first approximation

$$\mathrm{P} = 2\cdot\frac{\mathrm{A}^2}{8\pi}\cos^2\phi.$$

All problems in the optics of moving bodies can be solved by the method here employed. What is essential is, that the electric and magnetic force of the light which is influenced by a moving body, be transformed into a system of co-ordinates at rest relatively to the body. By this means all problems in the optics of moving bodies will be reduced to a series of problems in the optics of stationary bodies.

§ 9. Transformation of the Maxwell-Hertz Equations when Convection-Currents are Taken into Account

We start from the equations

$$\begin{aligned}
\frac{1}{c}\left\{\frac{\partial \mathrm{X}}{\partial t}+u_x\rho\right\} &= \frac{\partial \mathrm{N}}{\partial y}-\frac{\partial \mathrm{M}}{\partial z}, & \frac{1}{c}\frac{\partial \mathrm{L}}{\partial t} &= \frac{\partial \mathrm{Y}}{\partial z}-\frac{\partial \mathrm{Z}}{\partial y},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Y}}{\partial t}+u_y\rho\right\} &= \frac{\partial \mathrm{L}}{\partial z}-\frac{\partial \mathrm{N}}{\partial x}, & \frac{1}{c}\frac{\partial \mathrm{M}}{\partial t} &= \frac{\partial \mathrm{Z}}{\partial x}-\frac{\partial \mathrm{X}}{\partial z},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Z}}{\partial t}+u_z\rho\right\} &= \frac{\partial \mathrm{M}}{\partial x}-\frac{\partial \mathrm{L}}{\partial y}, & \frac{1}{c}\frac{\partial \mathrm{N}}{\partial t} &= \frac{\partial \mathrm{X}}{\partial y}-\frac{\partial \mathrm{Y}}{\partial x},
\end{aligned}$$

where

$$\rho=\frac{\partial \mathrm{X}}{\partial x}+\frac{\partial \mathrm{Y}}{\partial y}+\frac{\partial \mathrm{Z}}{\partial z}$$

denotes 4π times the density of electricity, and (u_x, u_y, u_z) the velocity-vector of the charge. If we imagine the electric charges to be invariably coupled to small rigid bodies (ions, electrons), these equations are the electromagnetic basis of the Lorentzian electrodynamics and optics of moving bodies.

Let these equations be valid in the system K, and transform them, with the assistance of the equations of transformation given in §§ 3 and 6, to the system k. We then obtain the equations

$$\begin{aligned}
\frac{1}{c}\left\{\frac{\partial \mathrm{X}'}{\partial \tau}+u_\xi\rho'\right\} &= \frac{\partial \mathrm{N}'}{\partial \eta}-\frac{\partial \mathrm{M}'}{\partial \zeta}, & \frac{1}{c}\frac{\partial \mathrm{L}'}{\partial \tau} &= \frac{\partial \mathrm{Y}'}{\partial \zeta}-\frac{\partial \mathrm{Z}'}{\partial \eta},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Y}'}{\partial \tau}+u_\eta\rho'\right\} &= \frac{\partial \mathrm{L}'}{\partial \zeta}-\frac{\partial \mathrm{N}'}{\partial \xi}, & \frac{1}{c}\frac{\partial \mathrm{M}'}{\partial \tau} &= \frac{\partial \mathrm{Z}'}{\partial \xi}-\frac{\partial \mathrm{X}'}{\partial \zeta},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Z}'}{\partial \tau}+u_\zeta\rho'\right\} &= \frac{\partial \mathrm{M}'}{\partial \xi}-\frac{\partial \mathrm{L}'}{\partial \eta}, & \frac{1}{c}\frac{\partial \mathrm{N}'}{\partial \tau} &= \frac{\partial \mathrm{X}'}{\partial \eta}-\frac{\partial \mathrm{Y}'}{\partial \xi},
\end{aligned}$$

where

$$\begin{aligned}
u_\xi &= \frac{u_x-v}{1-u_xv/c^2}\\
u_\eta &= \frac{u_y}{\beta(1-u_xv/c^2)}\\
u_\zeta &= \frac{u_z}{\beta(1-u_xv/c^2)},
\end{aligned}$$

and

$$\begin{aligned}
\rho' &= \frac{\partial \mathrm{X}'}{\partial \xi}+\frac{\partial \mathrm{Y}'}{\partial \eta}+\frac{\partial Z'}{\partial \zeta}\\
&= \beta(1-u_xv/c^2)\rho.
\end{aligned}$$

Since—as follows from the theorem of addition of velocities (§ 5)—the vector (u_ξ, u_η, u_ζ) is nothing else than the velocity of the electric charge, measured in the system k, we have the proof that, on the basis of our kinematical principles, the electrodynamic foundation of Lorentz's theory of the electrodynamics of moving bodies is in agreement with the principle of relativity.

In addition I may briefly remark that the following important law may easily be deduced from the developed equations: If an electrically charged body is in motion anywhere in space without altering its charge when regarded from a system of co-ordinates moving with the body, its charge also remains—when regarded from the "stationary" system K—constant.

§ 10. Dynamics of the Slowly Accelerated Electron

Let there be in motion in an electromagnetic field an electrically charged particle (in the sequel called an "electron"), for the law of motion of which we assume as follows:—

If the electron is at rest at a given epoch, the motion of the electron ensues in the next instant of time according to the equations

$$\begin{aligned} m\frac{d^2x}{dt^2} &= \epsilon\mathrm{X} \\ m\frac{d^2y}{dt^2} &= \epsilon\mathrm{Y} \\ m\frac{d^2z}{dt^2} &= \epsilon\mathrm{Z} \end{aligned}$$

where x, y, z denote the co-ordinates of the electron, and m the mass of the electron, as long as its motion is slow.

Now, secondly, let the velocity of the electron at a given epoch be v. We seek the law of motion of the electron in the immediately ensuing instants of time.

Without affecting the general character of our considerations, we may and will assume that the electron, at the moment when we give it our attention, is at the origin of the co-ordinates, and moves with the velocity v along the axis of X of the system K. It is then clear that at the given moment ($t = 0$) the electron is at rest relatively to a system of co-ordinates which is in parallel motion with velocity v along the axis of X.

From the above assumption, in combination with the principle of relativity, it is clear that in the immediately ensuing time (for small values of t) the electron, viewed from the system k, moves in accordance with the equations

$$\begin{aligned} m\frac{d^2\xi}{d\tau^2} &= \epsilon\mathrm{X}', \\ m\frac{d^2\eta}{d\tau^2} &= \epsilon\mathrm{Y}', \end{aligned}$$

$$m\frac{d^2\zeta}{d\tau^2} = \epsilon \mathrm{Z}',$$

in which the symbols ξ, η, ζ, X′, Y′, Z′ refer to the system k. If, further, we decide that when $t = x = y = z = 0$ then $\tau = \xi = \eta = \zeta = 0$, the transformation equations of §§ 3 and 6 hold good, so that we have

$$\xi = \beta(x - vt), \eta = y, \zeta = z, \tau = \beta(t - vx/c^2),$$
$$\mathrm{X}' = \mathrm{X}, \mathrm{Y}' = \beta(\mathrm{Y} - v\mathrm{N}/c), \mathrm{Z}' = \beta(\mathrm{Z} + v\mathrm{M}/c).$$

With the help of these equations we transform the above equations of motion from system k to system K, and obtain

$$\left.\begin{aligned} \frac{d^2x}{dt^2} &= \frac{\epsilon}{m\beta^3}\mathrm{X} \\ \frac{d^2y}{dt^2} &= \frac{\epsilon}{m\beta}\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right) \\ \frac{d^2z}{dt^2} &= \frac{\epsilon}{m\beta}\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right) \end{aligned}\right\} \quad . \quad . \quad . \quad (\mathrm{A})$$

Taking the ordinary point of view we now inquire as to the "longitudinal" and the "transverse" mass of the moving electron. We write the equations (A) in the form

$$\begin{aligned} m\beta^3\frac{d^2x}{dt^2} &= \epsilon\mathrm{X} &&= \epsilon\mathrm{X}', \\ m\beta^2\frac{d^2y}{dt^2} &= \epsilon\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right) &&= \epsilon\mathrm{Y}', \\ m\beta^2\frac{d^2z}{dt^2} &= \epsilon\beta\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right) &&= \epsilon\mathrm{Z}', \end{aligned}$$

and remark firstly that ϵX′, ϵY′, ϵZ′ are the components of the ponderomotive force acting upon the electron, and are so indeed as viewed in a system moving at the moment with the electron, with the same velocity as the electron. (This force might be measured, for example, by a spring balance at rest in the last-mentioned system.) Now if we call this force simply "the force acting upon the electron,"[9] and maintain the equation—mass × acceleration = force—and if we also decide that the accelerations are to be measured in the stationary system K, we derive from the above equations

$$\text{Longitudinal mass} = \frac{m}{(\sqrt{1 - v^2/c^2})^3}.$$
$$\text{Transverse mass} = \frac{m}{1 - v^2/c^2}.$$

With a different definition of force and acceleration we should naturally obtain other values for the masses. This shows us that in comparing different theories of the motion of the electron we must proceed very cautiously.

We remark that these results as to the mass are also valid for ponderable material points, because a ponderable material point can be made into an electron (in our sense of the word) by the addition of an electric charge, *no matter how small.*

[9]The definition of force here given is not advantageous, as was first shown by M. Planck. It is more to the point to define force in such a way that the laws of momentum and energy assume the simplest form.

We will now determine the kinetic energy of the electron. If an electron moves from rest at the origin of co-ordinates of the system K along the axis of X under the action of an electrostatic force X, it is clear that the energy withdrawn from the electrostatic field has the value $\int \epsilon \mathrm{X}\, dx$. As the electron is to be slowly accelerated, and consequently may not give off any energy in the form of radiation, the energy withdrawn from the electrostatic field must be put down as equal to the energy of motion W of the electron. Bearing in mind that during the whole process of motion which we are considering, the first of the equations (A) applies, we therefore obtain

$$\begin{aligned} \mathrm{W} &= \int \epsilon \mathrm{X}\, dx = m \int_0^v \beta^3 v\, dv \\ &= mc^2 \left\{ \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right\}. \end{aligned}$$

Thus, when $v = c$, W becomes infinite. Velocities greater than that of light have—as in our previous results—no possibility of existence.

This expression for the kinetic energy must also, by virtue of the argument stated above, apply to ponderable masses as well.

We will now enumerate the properties of the motion of the electron which result from the system of equations (A), and are accessible to experiment.

1. From the second equation of the system (A) it follows that an electric force Y and a magnetic force N have an equally strong deflective action on an electron moving with the velocity v, when $\mathrm{Y} = \mathrm{N}v/c$. Thus we see that it is possible by our theory to determine the velocity of the electron from the ratio of the magnetic power of deflexion A_m to the electric power of deflexion A_e, for any velocity, by applying the law

$$\frac{\mathrm{A}_m}{\mathrm{A}_e} = \frac{v}{c}.$$

This relationship may be tested experimentally, since the velocity of the electron can be directly measured, e.g. by means of rapidly oscillating electric and magnetic fields.

2. From the deduction for the kinetic energy of the electron it follows that between the potential difference, P, traversed and the acquired velocity v of the electron there must be the relationship

$$\mathrm{P} = \int \mathrm{X} dx = \frac{m}{\epsilon} c^2 \left\{ \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right\}.$$

3. We calculate the radius of curvature of the path of the electron when a magnetic force N is present (as the only deflective force), acting perpendicularly to the velocity of the electron. From the second of the equations (A) we obtain

$$-\frac{d^2y}{dt^2} = \frac{v^2}{\mathrm{R}} = \frac{\epsilon}{m}\frac{v}{c}\mathrm{N}\sqrt{1-\frac{v^2}{c^2}}$$

or

$$\mathrm{R} = \frac{mc^2}{\epsilon}\cdot\frac{v/c}{\sqrt{1-v^2/c^2}}\cdot\frac{1}{\mathrm{N}}.$$

These three relationships are a complete expression for the laws according to which, by the theory here advanced, the electron must move.

In conclusion I wish to say that in working at the problem here dealt with I have had the loyal assistance of my friend and colleague M. Besso, and that I am indebted to him for several valuable suggestions.

About this Document

This edition of Einstein's *On the Electrodynamics of Moving Bodies* is based on the English translation of his original 1905 German-language paper (published as *Zur Elektrodynamik bewegter Körper*, in *Annalen Der Physik*. **17**:891, 1905) which appeared in the book *The Principle of Relativity*, published in 1923 by Methuen and Company, Ltd. of London. Most of the papers in that collection are English translations from the German *Das Relativatsprinzip*, 4th ed., published by in 1922 by Tuebner.

Numbered footnotes are as they appeared in the 1923 edition; editor's notes are marked by a dagger (†) and appear in sans serif type. The 1923 English translation modified the notation used in Einstein's 1905 paper to conform to that in use by the 1920's; for example, c denotes the speed of light, as opposed the V used by Einstein in 1905.

This edition was prepared by John Walker. The current version of this document is available in a variety of formats from the editor's Web site:

`http://www.fourmilab.ch/`

DOES THE INERTIA OF A BODY DEPEND UPON ITS ENERGY-CONTENT?

By A. EINSTEIN

September 27, 1905

The results of the previous investigation lead to a very interesting conclusion, which is here to be deduced.

I based that investigation on the Maxwell-Hertz equations for empty space, together with the Maxwellian expression for the electromagnetic energy of space, and in addition the principle that:—

The laws by which the states of physical systems alter are independent of the alternative, to which of two systems of coordinates, in uniform motion of parallel translation relatively to each other, these alterations of state are referred (principle of relativity).

With these principles* as my basis I deduced *inter alia* the following result (§ 8):—

Let a system of plane waves of light, referred to the system of co-ordinates (x, y, z), possess the energy l; let the direction of the ray (the wave-normal) make an angle ϕ with the axis of x of the system. If we introduce a new system of co-ordinates (ξ, η, ζ) moving in uniform parallel translation with respect to the system (x, y, z), and having its origin of co-ordinates in motion along the axis of x with the velocity v, then this quantity of light—measured in the system (ξ, η, ζ)—possesses the energy

$$l^* = l\frac{1 - \frac{v}{c}cos\phi}{\sqrt{1 - v^2/c^2}}$$

where c denotes the velocity of light. We shall make use of this result in what follows.

Let there be a stationary body in the system (x, y, z), and let its energy—referred to the system (x, y, z) be E_0. Let the energy of the body relative to the system (ξ, η, ζ) moving as above with the velocity v, be H_0.

Let this body send out, in a direction making an angle ϕ with the axis of x, plane waves of light, of energy $\frac{1}{2}\mathrm{L}$ measured relatively to (x, y, z), and simultaneously an equal quantity of light in the opposite direction. Meanwhile the body remains at rest with respect to the system (x, y, z). The principle of

*The principle of the constancy of the velocity of light is of course contained in Maxwell's equations.

energy must apply to this process, and in fact (by the principle of relativity) with respect to both systems of co-ordinates. If we call the energy of the body after the emission of light E_1 or H_1 respectively, measured relatively to the system (x, y, z) or (ξ, η, ζ) respectively, then by employing the relation given above we obtain

$$\begin{aligned} E_0 &= E_1 + \frac{1}{2}L + \frac{1}{2}L, \\ H_0 &= H_1 + \frac{1}{2}L\frac{1 - \frac{v}{c}cos\phi}{\sqrt{1 - v^2/c^2}} + \frac{1}{2}L\frac{1 + \frac{v}{c}cos\phi}{\sqrt{1 - v^2/c^2}} \\ &= H_1 + \frac{L}{\sqrt{1 - v^2/c^2}}. \end{aligned}$$

By subtraction we obtain from these equations

$$H_0 - E_0 - (H_1 - E_1) = L\left\{\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right\}.$$

The two differences of the form $H - E$ occurring in this expression have simple physical significations. H and E are energy values of the same body referred to two systems of co-ordinates which are in motion relatively to each other, the body being at rest in one of the two systems (system (x, y, z)). Thus it is clear that the difference $H - E$ can differ from the kinetic energy K of the body, with respect to the other system (ξ, η, ζ), only by an additive constant C, which depends on the choice of the arbitrary additive constants of the energies H and E. Thus we may place

$$\begin{aligned} H_0 - E_0 &= K_0 + C, \\ H_1 - E_1 &= K_1 + C, \end{aligned}$$

since C does not change during the emission of light. So we have

$$K_0 - K_1 = L\left\{\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right\}.$$

The kinetic energy of the body with respect to (ξ, η, ζ) diminishes as a result of the emission of light, and the amount of diminution is independent of the properties of the body. Moreover, the difference $K_0 - K_1$, like the kinetic energy of the electron (§ 10), depends on the velocity.

Neglecting magnitudes of fourth and higher orders we may place

$$K_0 - K_1 = \frac{1}{2}\frac{L}{c^2}v^2.$$

From this equation it directly follows that:—

If a body gives off the energy L in the form of radiation, its mass diminishes by L/c^2. The fact that the energy withdrawn from the body becomes energy of radiation evidently makes no difference, so that we are led to the more general conclusion that

The mass of a body is a measure of its energy-content; if the energy changes by L, the mass changes in the same sense by $\mathrm{L}/9 \times 10^{20}$, the energy being measured in ergs, and the mass in grammes.

It is not impossible that with bodies whose energy-content is variable to a high degree (e.g. with radium salts) the theory may be successfully put to the test.

If the theory corresponds to the facts, radiation conveys inertia between the emitting and absorbing bodies.

About this Document

This edition of Einstein's *Does the Inertia of a Body Depend upon its Energy-Content* is based on the English translation of his original 1905 German-language paper (published as *Ist die Trägheit eines Körpers von seinem Energiegehalt abhängig?*, in *Annalen der Physik*. **18**:639, 1905) which appeared in the book *The Principle of Relativity*, published in 1923 by Methuen and Company, Ltd. of London. Most of the papers in that collection are English translations by W. Perrett and G.B. Jeffery from the German *Das Relativatsprinzip*, 4th ed., published by in 1922 by Tuebner. All of these sources are now in the public domain; this document, derived from them, remains in the public domain and may be reproduced in any manner or medium without permission, restriction, attribution, or compensation.

The footnote is as it appeared in the 1923 edition. The 1923 English translation modified the notation used in Einstein's 1905 paper to conform to that in use by the 1920's; for example, c denotes the speed of light, as opposed the V used by Einstein in 1905. In this paper Einstein uses L to denote energy; the italicised sentence in the conclusion may be written as the equation "$m = \mathrm{L}/c^2$" which, using the more modern E instead of L to denote energy, may be trivially rewritten as "$\mathrm{E} = mc^2$".

This edition was prepared by John Walker. The current version of this document is available in a variety of formats from the editor's Web site:

`http://www.fourmilab.ch/`

INTRODUCTION.

At the present time, different opinions are being held about the fundamental equations of Electro-dynamics for moving bodies. The Hertzian* forms must be given up, for it has appeared that they are contrary to many experimental results.

In 1895 H. A. Lorentz† published his theory of optical and electrical phenomena in moving bodies; this theory was based upon the atomistic conception (vorstellung) of electricity, and on account of its great success appears to have justified the bold hypotheses, by which it has been ushered into existence. In his theory, Lorentz proceeds from certain equations, which must hold at every point of "Äther"; then by forming the average values over "Physically infinitely small" regions, which however contain large numbers of electrons, the equations for electro-magnetic processes in moving bodies can be successfully built up.

In particular, Lorentz's theory gives a good account of the non-existence of relative motion of the earth and the luminiferous "Äther"; it shows that this fact is intimately connected with the covariance of the original equation, when certain simultaneous transformations of the space and time co-ordinates are effected; these transformations have therefore obtained from H. Poincare‡ the name of Lorentz-transformations. The covariance of these fundamental equations, when subjected to the Lorentz-transformation is a purely mathematical fact *i.e.* not based on any physical considerations; I will call this the Theorem of Relativity; this theorem rests essentially on the form of the

* *Vide* Note 1. † Note 2. ‡ *Vide* Note 3.

differential equations for the propagation of waves with the velocity of light.

Now without *recognizing* any hypothesis about the connection between " Äther " and matter, we can expect these mathematically evident theorems to have their consequences so far extended—that thereby even those laws of ponderable media which are yet unknown may anyhow possess this covariance when subjected to a Lorentz-transformation; by saying this, we do not indeed express an opinion, but rather a conviction,—and this conviction I may be permitted to call the Postulate of Relativity. The position of affairs here is almost the same as when the Principle of Conservation of Energy was poslutated in cases, where the corresponding forms of energy were unknown.

Now if hereafter, we succeed in maintaining this covariance as a definite connection between pure and simple observable phenomena in moving bodies, the definite connection may be styled 'the Principle of Relativity.'

These differentiations seem to me to be necessary for enabling us to characterise the present day position of the electro-dynamics for moving bodies.

H. A. Lorentz* has found out the " Relativity theorem" and has created the Relativity-postulate as a hypothesis that electrons and matter suffer contractions in consequence of their motion according to a certain law.

A. Einstein† has brought out the point very clearly, that this postulate is not an artificial hypothesis but is rather a new way of comprehending the time-concept which is forced upon us by observation of natural phenomena.

The Principle of Relativity has not yet been formulated for electro-dynamics of moving bodies in the sense

* *Vide* Note 4. † Note 5.

characterized by me. "In the present essay, while formulating this principle, I shall obtain the fundamental equations for moving bodies in a sense which is uniquely determined by this principle.

But it will be shown that none of the forms hitherto assumed for these equations can exactly fit in with this principle.*

We would at first expect that the fundamental equations which are assumed by Lorentz for moving bodies would correspond to the Relativity Principle. But it will be shown that this is not the case for the general equations which Lorentz has for any possible, and also for magnetic bodies; but this is approximately the case (if neglect the square of the velocity of matter in comparison to the velocity of light) for those equations which Lorentz hereafter infers for non-magnetic bodies. But this latter accordance with the Relativity Principle is due to the fact that the condition of non-magnetisation has been formulated in a way not corresponding to the Relativity Principle; therefore the accordance is due to the fortuitous compensation of two contradictions to the Relalivity-Postulate. But meanwhile enunciation of the Principle in a rigid manner does not signify any contradiction to the hypotheses of Lorentz's molecular theory, but it shall become clear that the assumption of the contraction of the electron in Lorentz's theory must be introduced' at an earlier stage than Lorentz has actually done.

In an appendix, I have gone into discussion of the position of Classical Mechanics with respect to the Relativity Postulate. Any easily perceivable modification of mechanics for satisfying the requirements of the Relativity theory would hardly afford any noticeable difference in observable processes; but would lead to very

* See notes on § 8 and 10.

surprising consequences. By laying down the Relativity-Postulate from the outset, sufficient means have been created for deducing henceforth the complete series of Laws of Mechanics from the principle of conservation of Energy alone (the form of the Energy being given in explicit forms).

NOTATIONS.

Let a rectangular system $(x, y, z, t,)$ of reference be given in space and time. The unit of time shall be chosen in such a manner with reference to the unit of length that the velocity of light in space becomes unity.

Although I would prefer not to change the notations used by Lorentz, it appears important to me to use a different selection of symbols, for thereby certain homogeneity will appear from the very beginning. I shall denote the vector electric force by E, the magnetic induction by M, the electric induction by e and the magnetic force by m, so that (E, M, e, m) are used instead of Lorentz's (E, B, D, H) respectively.

I shall further make use of complex magnitudes in a way which is not yet current in physical investigations, *i.e.*, instead of operating with (t), I shall operate with $(i\,t)$, where i denotes $\sqrt{-1}$. If now instead of $(x, y, z, i\,t)$, I use the method of writing with indices, certain essential circumstances will come into evidence ; on this will be based a general use of the suffixes (1, 2, 3, 4). The advantage of this method will be, as I expressly emphasize here, that we shall have to handle symbols which have apparently a purely real appearance; we can however at any moment pass to real equations if it is understood that of the symlbols with indices, such ones as have the suffix 4 only once, denote imaginary quantities, while those which have not at all the suffix 4, or have it twice denote real quantities.

An individual system of values of (x, y, z, t) *i. e.*, of $(x_1\ x_2\ r_3\ x_4)$ shall be called a space-time point.

Further let u denote the velocity vector of matter, ϵ the dielectric constant, μ the magnetic permeability, σ the conductivity of matter, while ρ denotes the density of electricity in space, and s the vector of "Electric Current" which we shall some across in §7 and §8.

PART I § 2.

THE LIMITING CASE.

The Fundamental Equations for Äther.

By using the electron theory, Lorentz in his above mentioned essay traces the Laws of Electro-dynamics of Ponderable Bodies to still simpler laws. Let us now adhere to these simpler laws, whereby we require that for the limitting case $\epsilon=1$, $\mu=1, \sigma=o$, they should constitute the laws for ponderable bodies. In this ideal limitting case $\epsilon=1$, $\mu=1$, $\sigma=o$, E will be equal to e, and M to m. At every space time point (x, y, z, t) we shall have the equations*

(i) $\text{Curl } m - \dfrac{\delta e}{\delta t} = \rho u$

(ii) $\text{div } e = \rho$

(iii) $\text{Curl } e + \dfrac{\delta m}{\delta t} = 0$

(iv) $\text{div m} = o$

I shall now write $(x_1\ x_2\ x_3\ x_4)$ for (x, y, z, t) and $(\rho_1, \rho_2, \rho_3, \rho_4)$ for

$(\rho u_x, \rho u_y, \rho u_z, i\rho)$

i.e. the components of the convection current ρu, and the electric density multiplied by $\sqrt{-1}$.

Further I shall write

$f_{23}, f_{31}, f_{12}, f_{14}, f_{24}, f_{34}.$

for

$m_x, m_y, m_z, -ie_x, -ie\ , -ie_z.$

i.e., the components of m and ($-i$.e.) along the three axes; now if we take any two indices (h. k) out of the series 3, 4), $f_{kh} = -f_{hk}$,

* See note 9

Therefore

$$f_{32} = -f_{23},\ f_{13} = -f_{31},\ f_{21} = -f_{12}$$
$$f_{41} = -f_{14},\ f_{44} = -f_{24},\ f_{43} = -f_{34}$$

Then the three equations comprised in (i), and the equation (ii) multiplied by i becomes

$$\left|\begin{array}{lllll} & \frac{\delta f_{12}}{\delta x_2} + & \frac{\delta f_{13}}{\delta x_3} + & \frac{\delta f_{14}}{\delta x_4} & = \rho_1 \\ \frac{\delta f_{21}}{\delta x_1} & + & \frac{\delta f_{23}}{\delta x_3} \times & \frac{\delta f_{24}}{\delta x_4} & = \rho_2 \\ \frac{\delta f_{31}}{\delta x_1} \times & \frac{\delta f_{32}}{\delta x_2} & + & \frac{\delta f_{34}}{\delta x_4} & = \rho_3 \\ \frac{\delta f_{41}}{\delta x_1} + & \frac{\delta f_{42}}{\delta x_2} + & \frac{\delta f_{43}}{\delta x_3} & & = \rho_4 \end{array}\right| \times \qquad \text{(A)}$$

On the other hand, the three equations comprised in (iii) and the (iv) equation multiplied by (i) becomes

$$\left|\begin{array}{lllll} & \frac{\delta f_{34}}{\delta x_2} + & \frac{\delta f_{42}}{\delta x_3} + & \frac{\delta f_{23}}{\delta x_4} & = 0 \\ \frac{\delta f_{43}}{\delta x_1} & + & \frac{\delta f_{14}}{\delta x_3} + & \frac{\delta f_{31}}{\delta x_4} & = 0 \\ \frac{\delta f_{24}}{\delta x_1} + & \frac{\delta f_{41}}{\delta x_2} & + & \frac{\delta f_{12}}{\delta x_4} & = 0 \\ \frac{\delta f_{32}}{\delta x_1} + & \frac{\delta f_{13}}{\delta x_2} + & \frac{\delta f_{21}}{\delta x_3} & & = 0 \end{array}\right| \times \qquad \text{(B)}$$

By means of this method of writing we at once notice the perfect symmetry of the 1st as well as the 2nd system of equations as regards permutation with the indices. (1, 2, 3, 4).

§ 3.

It is well-known that by writing the equations i) to iv) in the symbol of vector calculus, we at once set in evidence an invariance (or rather a (covariance) of the system of equations A) as well as of B), when the co-ordinate system is rotated through a certain amount round the null-point. For example, if we take a rotation of the axes round the z-axis. through an amount ϕ, keeping e, m fixed in space, and introduce new variables x_1', x_2' x_3' x_4' instead of x_1 x_2 x_3 x_4, where

$x'_1 = x_1 \cos\phi + x_2 \sin\phi, x'_2 = -x_1 \sin\phi + x_2 \cos\phi,$ $x'_3 = x_3 x'_4 = x_4$, and introduce magnitudes $\rho'_1, \rho'_2, \rho'_3 \rho'_4$, where $\rho_1' = \rho_1 \cos\phi + \rho_2 \sin\phi$, $\rho_2' = -\rho_1 \sin\phi + \rho_2 \cos\phi$ and f'_{12}, ... f'_{34}, where

$$f'_{23} = f_{23} \cos\phi + f_{31} \sin\phi, f'_{31} = -f_{23} \sin\phi + f_{31} \cos\phi, f'_{12} = f_{12},$$

$$f'_{14} = f_{14} \cos\phi + f_{24} \sin\phi, f'_{24} = -f_{14} \sin\phi + f_{24} \cos\phi, f'_{34} = f_{34},$$

$$f'_{kh} = -f_{kh} \text{ (h 1 k} = 1, 2, 3, 4).$$

then out of the equations (A) would follow a corresponding system of dashed equations (A′) composed of the newly introduced dashed magnitudes.

So upon the ground of symmetry alone of the equations (A) and (B) concerning the *suffixes* (1, 2, 3, 4), the theorem of Relativity, which was found out by Lorentz, follows without any calculation at all.

I will denote by $i\psi$ a purely imaginary magnitude, and consider the substitution

$$x_1' = x_1, x_2' = x_2, x_3' = x_3 \cos i\psi + x_4 \sin i\psi, \tag{1}$$

$$x_4' = -x_3 \sin i\psi + x_4 \cos i\psi,$$

$$\text{Putting } -i \tan i\psi = \frac{e^{\psi} - e^{-\psi}}{e^{\psi} + e^{-\psi}} = q, \psi = \frac{1}{2} \log \frac{1+q}{1-q}, \tag{2}$$

We shall have $\cos i\psi = \dfrac{1}{\sqrt{1-q^2}}$, $\sin i\psi = \dfrac{iq}{\sqrt{1-q^2}}$

where $-1 < q < 1$, and $\sqrt{1-q^2}$ is always to be taken with the positive sign.

Let us now write $x'_1 = x'$, $x'_2 = y'$, $x'_3 = z'$, $x'_4 = it'$ (3) then the substitution 1) takes the form

$$x' = x,\ y' = y,\ z' = \frac{z-qt}{\sqrt{1-q^2}},\ t' = \frac{-qz+t}{\sqrt{1-q^2}}, \qquad (4)$$

the coefficients being essentially real.

If now in the above-mentioned rotation round the Z-axis, we replace 1, 2, 3, 4 throughout by 3, 4, 1, 2, and ϕ by $i\psi$, we at once perceive that simultaneously, new magnitudes $\rho'_1, \rho'_2, \rho'_3, \rho'_4$, where

$(\rho'_1 = \rho_1, \rho'_2 = \rho_2, \rho'_3 = \rho_3 \cos i\psi + \rho_4 \sin i\psi, \rho'_4 =$
$-\rho_3 \sin i\psi + \rho_4 \cos i\psi)$,

and $f'_{12} \ldots f'_{34}$, where

$f'_{41} = f_{41} \cos i\psi + f_{13} \sin i\psi, f'_{13} = -f_{41} \sin i\psi + f_{13} \cos i\psi, f'_{34} = f_{34}, f'_{32} = f_{32} \cos i\psi + f_{42} \sin i\psi, f'_{42} = -f_{32} \sin i\psi + f_{42} \cos i\psi, f'_{12} = f_{12}, f_{kh} = -f'_{kh}$,

must be introduced. Then the systems of equations in (A) and (B) are transformed into equations (A′), and (B′), the new equations being obtained by simply dashing the old set.

All these equations can be written in purely real figures, and we can then formulate the last result as follows.

If the real transformations 4) are taken, and $x'\ y'\ z'\ t'$ be takes as a new frame of reference, then we shall have

$$(5)\quad \rho' = \rho\left[\frac{-qu_z+1}{\sqrt{1-q^2}}\right],\ \rho' u_z' = \rho\left[\frac{u_z-q}{\sqrt{1-q^2}}\right],$$
$$\rho' u_x' = \rho u_x,\ \rho' u_y' = \rho u_y.$$

$$(6)\ e_x{}' = \frac{e_x - qm_y}{\sqrt{1-q^2}},\ m'_y{}' = \frac{qe_x + m_y}{\sqrt{1-q^2}},\ e_z{}' = e_z.$$

$$(7)\ m'_x{}' = \frac{m_x + qe_y}{\sqrt{1-q^2}},\ e'_y{}' = \frac{qm_x + e_y}{\sqrt{1-q^2}},\ m'_z{}' = m_z.$$

Then we have for these newly introduced vectors u', e', m' (with components $u_x{}'$, $u_y{}'$, $u_z{}'$; $e_x{}'$, $e_y{}'$, $e_z{}'$; $m_x{}'$, $m_y{}'$, $m_z{}'$), and the quantity ρ' a series of equations I′), II′), III′), IV′) which are obtained from I), II), III), IV) by simply dashing the symbols.

We remark here that $e_x - qm_y$, $e_y + qm_x$ are components of the vector $e + [vm]$, where v is a vector in the direction of the positive Z-axis, and $|v| = q$, and $[vm]$ is the vector product of v and m; similarly $-qe_x + m_y$, $m_x + qe_y$ are the components of the vector $m - [ve]$.

The equations 6) and 7), as they stand in pairs, can be expressed as.

$$e_x{}' + im'_x{}' = (e_x + im_x) \cos i\psi + (e_y + im_y) \sin i\psi,$$

$$e_y{}' + im'_y{}' = -(e_x + im_x) \sin i\psi + (e_y + im_y) \cos i\psi,$$

$$e_z{}' + im'_z{}' = e'_z + im_z.$$

If ϕ denotes any other real angle, we can form the following combinations :—

$$(e_x{}' + im'_x{}') \cos.\ \phi + (e_y{}'' + im'_y{}') \sin \phi$$

$$= (e_x + im_x) \cos.\ (\phi + i\psi) + (e_y + im_y) \sin (\phi + i\psi),$$

$$= (e_x{}' + im'_x{}') \sin \phi + (e_y{}' + im'_y{}') \cos.\ \phi$$

$$= -(e_x + im_x) \sin (\phi + i\psi) + (e_y + im_y) \cos.\ (\phi + i\psi).$$

Special Lorentz Transformation.

The rôle which is played by the Z-axis in the transformation (4) can easily be transferred to any other axis when the system of axes are subjected to a transformation about this last axis. So we came to a more general law :—

Let v be a vector with the components v_x, v_y, v_z, and let $|v| = q < 1$. By $\bar{v}$ we shall denote any vector which is perpendicular to v, and by r_v, $r_{\bar{v}}$ we shall denote components of r in direction of $\bar{v}$ and v.

Instead of (x, y, z, t), new magnetudes $(x'\ y'\ z'\ t')$ will be introduced in the following way. If for the sake of shortness, r is written for the vector with the components (x, y, z) in the first system of reference, r' for the same vector with the components $(x'\ y'\ z')$ in the second system of reference, then for the direction of v, we have

$$(10) \quad r'_v = \frac{r_v - qt}{\sqrt{1-q^2}}$$

and for the perpendicular direction $\bar{v}$,

$$(11) \quad r'_{\bar{v}} = r_{\bar{v}}$$

and further (12) $t' = \dfrac{-qr_v + t}{\sqrt{1-q^2}}$.

The notations $(r'_{\bar{v}}, r'_v)$ are to be understood in the sense that with the directions v, and every direction $\bar{v}$ perpendicular to v in the system (x, y, z) are always associated the directions with the same direction cosines in the system $(x'\ y, z')$.

A transformation which is accomplished by means of (10), (11), (12) with the condition $0 < q < 1$ will be called a special Lorentz-transformation. We shall call v the vector, the direction of v the axis, and the magnitude of v the moment of this transformation.

If further ρ' and the vectors u', e', m', in the system $(x'y'z')$ are so defined that,

$$(13) \quad \rho' = \rho\left[\frac{-qu_v + 1}{\sqrt{1-q^2}}\right],\ \rho' u'_v = \frac{\rho(u_v - q)}{\sqrt{1-q^2}},\ \rho' u'_{\bar{v}} =,\ \rho' u_v,$$

further

(14) $(e'+im')_{\bar{v}} = \dfrac{(e+im) - i[v, (e+im])_{\bar{v}}}{\sqrt{1-q^2}}.$

(15) $(e'+im')_v = (e+im) - i\ [u, (e+im)]_v.$

Then it follows that the equations I), II), III), IV) are transformed into the corresponding system with dashes.

The solution of the equations (10), (11), (12) leads to

(16) $r_v = \dfrac{r'_v + qt'}{\sqrt{1-q^2}}, \ r_{\bar{v}}, = r'_{\bar{v}}, \ t = \dfrac{qr'_v + t'}{\sqrt{1-q^2}},$

Now we shall make a very important observation about the vectors u and u'. We can again introduce the indices 1, 2, 3, 4, so that we write (x_1', x_2', x_3', x_4') instead of (x', y', z', it') and $\rho_1', \rho_2', \rho_3', \rho_4'$ instead of $(\rho' u'_x{}', \rho' u'_y{}', \rho' u'_z{}', i\rho'$.

Like the rotation round the Z-axis, the transformation (4), and more generally the transformations (10), (11), (12), are also linear transformations with the determinant $+1$, so that

(17) $x_1{}^2 + x_2{}^2 + x_3{}^2 + x_4{}^2$ *i. e.* $x^2 + y^2 + z^2 - t^2$,

is transformed into

$x_1'^2 + x_2'^2 + x_3'^2 + x_4'^2$ *i. e.* $x'^2 + y'^2 + z'^2 - t'^2$.

On the basis of the equations (13), (14), we shall have $(\rho_1{}^2 + \rho_2{}^2 + \rho_3{}^2 + \rho_4{}^2) = \rho^2(1 - u_x{}^2, - u_y{}^2, - u_z{}^2,) = \rho^2(1-u^2)$ transformed into $\rho^2(1-u^2)$ or in other words,

(18) $\rho\sqrt{1-u^2}$

is an invariant in a Lorentz-transformation.

If we divide $(\rho_1, \rho_2, \rho_3, \rho_4)$ by this magnitude, we obtain the four values $(\omega_1, \omega_2, \omega_3, \omega_4) = \dfrac{1}{\sqrt{1-u^2}}(u_x, u_y, u_z, i)$

so that $\omega_1{}^2 + \omega_2{}^2 + \omega_3{}^2 + \omega_4{}^2 = -1$.

It is apparent that these four values, are determined by the vector u and inversely the vector u of magnitude

$<$ follows from the 4 values $\omega_1, \omega_2, \omega_3, \omega_4$; where $(\omega_1, \omega_2, \omega_3)$ are real, $-i\omega_4$ real and positive and condition (19) is fulfilled.

The meaning of $(\omega_1, \omega_2, \omega_3, \omega_2)$ here is, that they are the ratios of $dx_1, dx_2, d\cdot_3, dx_4$ to

$$(20) \quad \sqrt{-(dx_1{}^2+dx_2{}^2+dx_3{}^2+dx_4{}^2}=dt\sqrt{1-u^2},$$

The differentials donoting the displacements of matter occupying the spacetime point (x_1, x_2, x_3, x_4) to the adjacent space-time point.

After the Lorentz-transfornation is accomplished the vococity of matter in the new system of reference for the same space-time point $(x' \; y' \; z' \; t')$ is the vector u' with the ratios $\frac{dx'}{dt'}, \frac{dy'}{dt'}, \frac{dz'}{dt'}, \frac{dl'}{dt'}$, as components.

Now it is quite apparent that the system of values

$$x_1=\omega_1, \; x_2=\omega_2, \; x_3=\omega_3, \; x_4=\omega_4$$

is transformed into the values

$$x_1'=\omega_1', \; x_2'=\omega_2', \; x_3'=\omega_3', \; x_4'=\omega_4'$$

in virtue of the Lorentz-transformation (10), (11), (12).

The dashed system has got the same meaning for the velocity u' after the transformation as the first system of values has got for u before transformation.

If in particular the vector v of the special Lorentz-transformation be equal to the velecity vector u of matter at the space-time point (x_1, x_2, x_3, x_4) then it follows out of (10), (11), (12) that

$$\omega_1'=o, \; \omega_2'=o, \; \omega_3'=o, \; \omega_4'=i$$

Under these circumstances therefore, the corresponding space-time point has the velocity $v'=o$ after the transformation, it is as if we transform to rest. We may call the invariant $\rho\sqrt{1-u^2}$ as the rest-density of Electricity.*

* See Note.

§ 5. Space-time Vectors.

Of the 1st and 2nd kind.

If we take the principal result of the Lorentz transformation together with the fact that the system (A) as well as the system (B) is covariant with respect to a rotation of the coordinate-system round the null point, we obtain the general *relativity theorem.* In order to make the facts easily comprehensible, it may be more convenient to define a series of expressions, for the purpose of expressing the ideas in a concise form, while on the other hand I shall adhere to the practice of using complex magnitudes, in order to render certain symmetries quite evident.

Let us take a linear homogeneous transformation,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{4} \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \end{bmatrix}$$

the Determinant of the matrix is $+1$, all co-efficients without the index 4 occurring once are real, while a_{41}, a_{42}, a_{43}, are purely imaginary, but a_{44} is real and $>o$, and $x_1{}^2+x_2{}^2+x_3{}^2+x_4{}^2$ transforms into $x_1'^2+x_2'^2+x_3'^2+x_4'^2$. The operation shall be called a general Lorentz transformation.

If we put $x_1'=x'$, $x_2'=y'$, $x_3'=z'$, $x_4'=it'$, then immediately there occurs a homogeneous linear transformation of (x, y, z, t) to (x', y', z', t') with essentially real co-efficients, whereby the aggregrate $-x^2-y^2-z^2+t^2$ transforms into $-x'^2-y'^2-z'^2+t'^2$, and to every such system of values x, y, z, t with a positive t, for which this aggregate $>o$, there always corresponds a positive t';

This notation, which is due to Dr. C. E. Cullis of the Calcutta University, has been used throughout instead of Minkowski's notation, $x_1=a_{11}x_1'+a_{12}x_2'+a_{13}x_3'+a_{14}x_4'$.

this last is quite evident from the continuity of the aggregate x, y, z, t.

The last vertical column of co-efficients has to fulfil, the condition 22) $a_{14}^2+a_{24}^2+a_{34}^2+a_{44}^2=1$.

If $a_{14}=a_{24}=a_{34}=o$, then $a_{44}=1$, and the Lorentz transformation reduces to a simple rotation of the spatial co-ordinate system round the world-point.

If a_{14}, a_{24}, a_{34} are not all zoro, and if we put $a_{14} : a_{24} : a_{34} : a_{44}=v_x : v_y : v_z : i$

$$q=\sqrt{v_x^2+v_y^2+v_z^2}<1.$$

On the other hand, with every set of value of $a_{14}, a_{24}, a_{34}, a_{44}$ which in this way fulfil the condition 22) with real values of v_x, v_y, v_z, we can construct the special Lorentz transformation (16) with $(a_{14}, a_{24}, a_{34}, a_{44})$ as the last vertical column,—and then every Lorentz-transformation with the same last vertical column $(a_{14}, a_{24}, a_{34}, a_{44})$ can be supposed to be composed of the special Lorentz-transformation, and a rotation of the spatial co-ordinate system round the null-point.

The totality of all Lorentz-Transformations forms a group. Under a space-time vector of the 1st kind shall be understood a system of four magnitudes $\rho_1, \rho_2, \rho_3, \rho_4)$ with the condition that in case of a Lorentz-transformation it is to be replaced by the set $\rho_1', \rho_2', \rho_3', \rho_4')$, where these are the values of $x_1', x_2', x_3', x_4')$, obtained by substituting $(\rho_1, \rho_2, \rho_3, \rho\ )$ for (x_1, x_2, x_3, x_4) in the expression (21).

Besides the time-space vector of the 1st kind (x_1, x_2, x_3, x_4) we shall also make use of another space-time vector of the first kind (y_1, y_2, y_3, y_4), and let us form the linear combination

(23) $f_{23}(x_2 y_3-x_3 y_2)+f_{31}(x_3 y_1-x_1 y_3)+f_{12}(x_1 y_2-x_2 y_1)+f_{14}(x_1 y_4-x_4 y_1)+f_{24}(x_2 y_4-x_4 y_2)+f_{34}(x_3 y_4-x_4 y_3)$

with six coefficients $f_{23}-f_{34}$. Let us remark that in the vectorial method of writing, this can be constructed out of the four vectors.

x_1, x_2, x_3 ; y_1, y_2, y_3 ; f_{23}, f_{31}, f_{12} ; f_{14}, f_{24}, f_{34} and the constants x_4 and y_4, at the same time it is symmetrical with regard the indices (1, 2, 3, 4).

If we subject (x_1, x_2, x_3, x_4) and (y_1, y_2, y_3, y_4) simultaneously to the Lorentz transformation (21), the combination (23) is changed to.

(24) $f_{23}'\,(x_2'\,y_3'-x_3'\,y_2) + f_{31}\,(x_3'\;y_1'-x_1'\,y_3) + f_{12}\,(x_1'\,y_2'-x_2'\,y_1') + f_{14}'\,(x_1'\,y_4') - x_4'\,y_1') + f_{24}'\,(x_2'\,y_4' - x_4'\,y_2') + f_{34}'\,(x_3'\,y_4'-x_4'\;y_3')$,

where the coefficients $f_{23}', f_{31}', f_{12}', f_{14}', f_{24}', f_{34}'$, depend solely on $(f_{23}\,f_{24})$ and the coefficients $a_{11} \ldots a_{44}$.

We shall define a space-time Vector of the 2nd kind as a system of six-magnitudes $f_{23}, f_{31} \ldots\ldots f_{34}$, with the condition that when subjected to a Lorentz transformation, it is changed to a new system $f_{23}' \ldots\ldots f_{34}', \ldots$ which satisfies the connection between (23) and (24).

I enunciate in the following manner the general theorem of relativity corresponding to the equations (I)—(iv),—which are the fundamental equations for Äther.

If x, y, z, it (space co-ordinates, and time it) is subjected to a Lorentz transformation, and at the same time $(\rho u_x, \rho u_y, \rho u_z, i\rho)$ (convection-current, and charge density ρi) is transformed as a space time vector of the 1st kind, further $(m_x, m_y, m_z, -ie_x, -ie_y, -ie_z)$ (magnetic force, and electric induction $\times$ $(-i)$ is transformed as a space time vector of the 2nd kind, then the system of equations (1), (II), and the system of equations (III), (IV) transforms into essentially corresponding relations between the corresponding magnitudes newly introduced info the system.

These facts can be more concisely expressed in these words: the system of equations (I, and II) as well as the system of equations (III) (IV) are covariant in all cases of Lorentz-transformation, where $(\rho u, i\rho)$ is to be transformed as a space time vector of the 1st kind, $(m-ie)$ is to be treated as a vector of the 2nd kind, or more significantly,—

$(\rho u, i\rho)$ is a space time vector of the 1st kind, $(m-ie)$* is a space-time vector of the 2nd kind.

I shall add a few more remarks here in order to elucidate the conception of space-time vector of the 2nd kind. Clearly, the following are invariants for such a vector when subjected to a group of Lorentz transformation.

$$(i) \quad m^2-e^2 = f_{23}^2 + f_{31}^2 + f_{12}^2 + f_{14}^2 + f_{24}^2 + f_{24}^2$$

$$me = i\,(f_{23}\,f_{14} + f_{31}\,f_{24} + f_{12}\,f_{34}).$$

A space-time vector of the second kind $(m-ie)$, where (m, and e) are real magnitudes, may be called singular, when the scalar square $(m-ie)^2=o$, ie $m^2-e^2=o$, and at the same time $(m\ e)=o$, ie the vector m and e are equal and perpendicular to each other; when such is the case, these two properties remain conserved for the space-time vector of the 2nd kind in every Lorentz-transformation.

If the space-time vector of the 2nd kind is not singular, we rotate the spacial co-ordinate system in such a manner that the vector-product $[me]$ coincides with the Z-axis, *i.e.* $m_z, = o$, $e_z = o$. Then

$$(m_x, -i\,e_x)^2 + (m_y, -i\,e_y)^2 \neq o,$$

Therefore $(e_y + i\,m_y,)/(e_x + i\,e_x)$ is different from $+i$, and we can therefore define a complex argument $\phi + i\psi)$ in such a manner that

$$\tan(\phi + i\,\psi) = \frac{e_y + i\,m_y}{e_x + i\,m_x},$$

* Vide Note.

If then, by referring back to equations (9), we carry out the transformation (1) through the angle ψ, and a subsequent rotation round the Z-axis through the angle ϕ, we perform a Lorentz-transformation at the end of which $m_y = o$, $e_y = o$, and therefore m and e shall both coincide with the new Z-axis. Then by means of the invariants $m^2 - e^2$, (me) the final values of these vectors, whether they are of the same or of opposite directions, or whether one of them is equal to zero, would be at once settled.

§ Concept of Time.

By the Lorentz transformation, we are allowed to effect certain *changes* of the time parameter. In consequence of this fact, it is no longer permissible to speak of the absolute simultaneity of two events. The ordinary idea of simultaneity rather presupposes that six independent parameters, which are evidently required for defining a system of space and time axes, are somehow reduced to three. Since we are accustomed to consider that these limitations represent in a unique way the actual facts very approximately, we maintain that the simultaneity of two events exists of themselves.* In fact, the following considerations will prove conclusive.

Let a reference system (x, y, z, t) for space time points (events) be somehow known. Now if a space point A (x_o, y_o, z_o) at the time t_o be compared with a space point P (x, y, z) at the time t, and if the difference of time $t - t_o$, (let $t > t_o$) be less than the length A P *i.e.* less than the time required for the propogation of light from

* Just as beings which are confined within a narrow region surrounding a point on a shperical surface, may fall into the error that a sphere is a geometric figure in which one diameter is particularly distinguished from the rest.

A to P, and if $q = \frac{t - t_o}{A\,P} < 1$, then by a special Lorentz transformation, in which A P is taken as the axis, and which has the moment q, we can introduce a time parameter t', which (see equation 11, 12, § 4) has got the same value $t' = o$ for both space-time points (A, t_o), and P, t). So the two events can now be comprehended to be simultaneous.

Further, let us take at the same time $t_o = o$, two different space-points A, B, or three space-points (A, B, C) which are not in the same space-line, and compare therewith a space point P, which is outside the line A B, or the plane A B C, at another time t, and let the time difference $t - t_o$ (t > t_o) be less than the time which light requires for propogation from the line A B, or the plane A B C) to P. Let q be the quotient of $(t - to)$ by the second time. Then if a Lorentz transformation is taken in which the perpendicular from P on A B, or from P on the plane A B C is the axis, and q is the moment, then all the three (or four) events (A, t_o), [B, t_o), (C, t_o) and (P, t) are simultaneous.

If four space-points. which do not lie in one plane are conceived to be at the same time t_o, then it is no longer permissible to make a change of the time parameter by a Lorentz—transformation, without at the same time destroying the character of the simultaneity of these four space points.

To the mathematician, accustomed on the one hand to the methods of treatment of the poly-dimensional manifold, and on the other hand to the conceptual figures of the so-called non-Euclidean Geometry, there can be no difficulty in adopting this concept of time to the application of the Lorentz-transformation. The paper of Einstein which has been cited in the Introduction, has succeeded to some extent in presenting the nature of the transformation from the physical standpoint.

PART II. ELECTRO-MAGNETIC PHENOMENA.

§ 7. Fundamental Equations for bodies at rest.

After these preparatory works, which have been first developed on account of the small amount of mathematics involved in the limitting case $\epsilon = 1, \mu = 1, \sigma = o$, let us turn to the electro-magnatic phenomena in matter. We look for those relations which make it possible for us —— when proper fundamental data are given — to obtain the following quantities at every place and time, and therefore at every space-time point as functions of (x, y, z, t):—the vector of the electric force E, the magnetic induction M, the electrical induction e, the magnetic force m, the electrical space-density ρ, the electric current s (whose relation hereafter to the conduction current is known by the manner in which conductivity occurs in the process), and lastly the vector u, the velocity of matter.

The relations in question can be divided into two classes.

Firstly—those equations, which,—when v, the velocity of matter is given as a function of (x, y, z, t),—lead us to a knowledge of other magnitude as functions of x, y, z, t —I shall call this first class of equations the fundamental equations—

Secondly, the expressions for the ponderomotive force, which, by the application of the Laws of Mechanics, gives us further information about the vector u as functions of x, y, z, t).

For the case of bodies at rest, *i.e.* when $u\ (x, y, z, t) = o$ the theories of Maxwell (Heaviside, Hertz) and

Lorentz lead to the same fundamental equations. They are ;—

(1) The Differential Equations :—which contain no constant referring to matter :—

$$(i)\ \text{Curl}\ m - \frac{\delta e}{\delta t} = C,\ (ii)\ \text{div}\ e = \rho.$$

$$(iii)\ \text{Curl}\ E + \frac{\delta M}{\delta t} = o,\ (iv)\ \text{Div}\ M = o.$$

(2) Further relations, which characterise the infiuence of existing matter for the most important case to which we limit ourselves *i.e.* for isotopic bodies ;—they are comprised in the equations

$$(V)\ e = \epsilon E,\ M = \mu m,\ C = \sigma E.$$

where ϵ = dielectric constant, μ = magnetic permeability, σ = the conductivity of matter, all given as function of x, y, z, t; s is here the conduction current.

By employing a modified form of writing, I shall now cause a latent symmetry in these equations to appear. I put, as in the previous work,

$$x_1 = x,\ x_2 = y,\ x_3 = z,\ x_4 = it,$$

and write s_1, s_2, s_3, s_4 for $C_x, C_y, C_z\ \sqrt{-1}\,\rho$.

further $f_{23}, f_{31}, f_{12}, f_{14}, f_{24}, f_{34}$

for $m_x, m_y, m_z - i\ (e_x, e_y, e_z)$,

and $F_{23}, F_{31}, F_{12}, F_{14}, F_{24}, F_{34}$

for $M_x, M_y, M_z, - i\ (E_x, E_y, E_z)$

lastly we shall have the relation $f_{kh} = -, f_{hk}$, $F_{kh} = -F_{hk}$, (the letter f, F shall denote the field, s the (*i.e.* current).

Then the fundamental Equations can be written as

$$
(A)\quad \left.\begin{array}{rcccl}
 & \dfrac{\partial f_{12}}{\partial x_2} + & \dfrac{\partial f_{13}}{\partial x_3} + & \dfrac{\partial f_{14}}{\partial x_4} & = s_1 \\[2ex]
\dfrac{\partial f_{21}}{\partial x_1} + & + & \dfrac{\partial f_{23}}{\partial x_3} + & \dfrac{\partial f_{24}}{\partial x_4} & = s_2 \\[2ex]
\dfrac{\partial f_{31}}{\partial x_1} + & \dfrac{\partial f_{32}}{\partial x_2} + & + & \dfrac{\partial f_{34}}{\partial x_4} & = s_3 \\[2ex]
\dfrac{\partial f_{41}}{\partial x_1} + & \dfrac{\partial f_{42}}{\partial x_2} + & \dfrac{\partial f_{43}}{\partial x_4} & & = s_4
\end{array}\right\}
$$

and the equations (3) and (4), are

$$
\left.\begin{array}{rcccl}
 & \dfrac{\partial F_{34}}{\partial x_2} + & \dfrac{\partial F_{42}}{\partial x_3} + & \dfrac{\partial F_{23}}{\partial x_4} & = 0 \\[2ex]
\dfrac{\partial F_{43}}{\partial x_1} & + & \dfrac{\partial F_{14}}{\partial x_3} + & \dfrac{\partial F_{31}}{\partial x_4} & = 0 \\[2ex]
\dfrac{\partial F_{24}}{\partial x_1} + & \dfrac{\partial F_{41}}{\partial x_2} + & + & \dfrac{\partial F_{12}}{\partial x_4} & = 0 \\[2ex]
\dfrac{\partial F_{32}}{\partial x_1} + & \dfrac{\partial F_{13}}{\partial x_2} + & \dfrac{\partial F_{21}}{\partial x_3} & & = 0
\end{array}\right\}
$$

§ 8. The Fundamental Equations.

We are now in a position to establish in a unique way the fundamental equations for bodies moving in any manner by means of these three axioms exclusively.

The first Axion shall be,—

When a detached region* of matter is at rest at any moment, therefore the vector u is zero, for a system

* Einzelne stelle der Materie.

(x, y, z, t)—the neighbourhood may be supposed to be in motion in any possible manner, then for the space-time point x, y, z, t, the same relations (A) (B) (V) which hold in the case when all matter is at rest, shall also hold between ρ, the vectors C, e, m, M, E and their differentials with respect to x, y, z, t. The second axiom shall be :—

Every velocity of matter is <1, smaller than the velocity of propogation of light.*

The fundamental equations are of such a kind that when (x, y, z, it) are subjected to a Lorentz transformation and thereby $(m-ie)$ and $(M-iE)$ are transformed into space-time vectors of the second kind, $(C, i\rho)$ as a space-time vector of the 1st kind, the equations are transformed into essentially identical forms involving the transformed magnitudes.

Shortly I can signify the third axiom as :—

$(m, -ie)$, and $(M, -iE)$ are space-time vectors of the second kind, $(C, i\rho)$ is a space-time vector of the first kind.

This axiom I call the Principle of Relativity.

In fact these three axioms lead us from the previously mentioned fundamental equations for bodies at rest to the equations for moving bodies in an unambignous way.

According to the second axiom, the magnitude of the velocity vector $|u|$ is <1 at any space-time point. In consequence, we can always write, instead of the vector u, the following set of four allied quantities

$$\omega_1 = \frac{u_x}{\sqrt{1-u^2}}, \quad \omega_2 = \frac{u_y}{\sqrt{1-u^2}}, \quad \omega_3 = \frac{u_z}{\sqrt{1-u^2}},$$

$$\omega_4 = \frac{i}{\sqrt{1-u^2}}$$

* Vide Note.

with the relation

(27) $\omega_1^2+\omega_2^2+\omega_3^2+\omega_4^2=-1$

From what has been said at the end of § 4, it is clear that in the case of a Lorentz-transformation, this set behaves like a space-time vector of the 1st kind.

Let us now fix our attention on a certain point (x, y, z) of matter at a certain time (t). If at this space-time point $u=o$, then we have at once for this point the equations (A), (B) (V) of § 7. It $u \ddagger o$, then there exists according to 16), in case $|u|<1$, a special Lorentz-transformation, whose vector v is equal to this vector u (x, y, z, t), and we pass on to a new system of reference $(x'\ y'\ z'\ t')$ in accordance with this transformation. Therefore for the space-time point considered, there arises as in § 4, the new values 28) $\omega'_1=O$, $\omega'_2=O$, $\omega'_3=O$, $\omega'_4=i$, therefore the new velocity vector $\omega'=o$, the space-time point is as if transformed to rest. Now according to the third axiom the system of equations for the transformed point $(x'\ y'\ z'\ t)$ involves the newly introduced magnitude $(u'\ \rho', C', e', m', E', M')$ and their differential quotients with respect to (x', y', z', t') in the same manner as the original equations for the point (x, y, z, t). But according to the first axiom, when $u'=o$, these equations must be exactly equivalent to

(1) the differential equations (A'), (B'), which are obtained from the equations (A), (B) by simply dashing the symbols in (A) and (B).

(2) and the equations

(V') $e'=\epsilon E'$, $M'=\mu m'$, $C'=\sigma E'$

where ϵ, μ, σ are the dielectric constant, magnetic permeability, and conductivity for the system $(x'\ y'\ z'\ t')$ *i.e.* in the space-time point $(x\ y, z\ t)$ of matter.

Now let us return, by means of the reciprocal Lorentz-transformation to the original variables (x, y, z, t), and the magnitudes (u, ρ, C, e, m, E, M) and the equations, which we then obtain from the last mentioned, will be the fundamental equations sought by us for the moving bodies.

Now from § 4, and § 6, it is to be seen that the equations $A)$, as well as the equations $B)$ are covariant for a Lorentz-transformation, *i.e.* the equations, which we obtain backwards from $A')$ $B')$, must be exactly of the same form as the equations $A)$ and $B)$, as we take them for bodies at rest. We have therefore as the first result :—

The differential equations expressing the fundamental equations of electrodynamics for moving bodies, when written in ρ and the vectors C, e, m, E, M, are exactly of the same form as the equations for moving bodies. The velocity of matter does not enter in these equations. In the vectorial way of writing, we have

$$\text{I}\Big)\ \text{curl}\ m - \frac{\partial e}{\partial t} = C_1, \qquad \text{II}\Big)\ \text{div}\ e = \rho$$

$$\text{III}\Big)\ \text{curl}\ E + \frac{\partial M}{\partial t} = 0 \qquad \text{IV}\Big)\ \text{div}\ M = 0$$

The velocity of matter occurs only in the auxilliary equations which characterise the influence of matter on the basis of their characteristic constants ϵ, μ, σ. Let us now transform these auxilliary equations $V')$ into the original co-ordinates $(x, y, z,$ and $t.)$

According to formula 15) in § 4, the component of e' in the direction of the vector u is the same us that of $(e + [u\ m])$, the component of m' is the same as that of $m - [u\ e]$, but for the perpendicular direction $\bar{u}$, the components of e', m' are the same as those of $(e + [u\ m])$ and $(m - [u\ e]$, multiplied by $\frac{1}{\sqrt{1-u^2}}$. On the other hand E' and M′ shall stand to $E + [uM,]$, and $M—[uE]$ in the same relation as e' and m' to $e+[um]$, and $m-(ue)$. From the relation $e'=\epsilon E'$, the following equations follow

$$(C) \quad e+[u\,m]=\epsilon(E+[u\,M]),$$

and from the relation $M'=\mu\, m'$, we have

$$(D) \quad M-[u\,E]=\mu\,(m-[u\,e]),$$

For the components in the directions perpendicular to u, and to each other, the equations are to be multiplied by $\sqrt{1-u^2}$.

Then the following equations follow from the transfermation equations (12), 10), (11) in § 4, when we replace q, r_v, $r_{\bar{v}}$, t, r'_v, $r'_{\bar{v}}$, t' by $|u|$, C_u, $C_{\bar{u}}$, ρ, C'_u $C'_{\bar{u}}$, ρ'

$$\rho'=\frac{-|u|C_u+\rho}{\sqrt{1-u^2}},\ C'_u=\frac{C_u-|u|\rho}{\sqrt{1-u^2}},\ C'_{\bar{u}}=C_{\bar{u}},$$

$$E\Big)\ \frac{C_u-|u|\rho}{\sqrt{1-u^2}}=\sigma\,(E+[u\,M])_u,$$

$$C_{\bar{u}}=\sigma\frac{(E+[uM])_u}{\sqrt{1-u^2}}$$

In consideration of the manner in which σ enters into these relations, it will be convenient to call the vector $C-\rho\, u$ with the components $C_v-\rho\,|u|$ in the direction of u, and C'_u in the directions $\bar{u}$ perpendicular to u the "Convection current." This last vanishes for $\sigma=0$.

We remark that for $\epsilon=1$, $\mu=1$ the equations $e'=E'$, $m'=M'$ immediately lead to the equations $e=E$, $m=M$ by means of a reciprocal Lorentz-transformation with—u as vector; and for $\sigma=0$, the equation $C'=0$ leads to $C=\rho\, u$; that the fundamental equations of Äther discussed in § 2 becomes in fact the limitting case of the equations obtained here with $\epsilon=1$, $\mu=1$, $\sigma=0$.

§9. The Fundamental Equations in Lorentz's Theory.

Let us now see how far the fundamental equations assumed by Lorentz correspond to the Relativity postulate, as defined in §8. In the article on Electron-theory (Ency, Math., Wiss., Bd. V. 2, Art 14) Lorentz has given the fundamental equations for any possible, even magnetised bodies (see there page 209, Eqⁿ XXX′, formula (14) on page 78 of the same (part).

(IIIa'') $\text{Curl}\ (H-[uE]) = J + \frac{dD}{dt} + u\ \text{div}\ D$

$-\text{curl}\ [uD].$

(I″) $\text{div}\ D = \rho$

(IV″) $\text{curl}\ E = -\frac{dB}{dt}$, $\text{Div}\ B = O$ (V′)

Then for moving non-magnetised bodies, Lorentz puts (page 223, 3) $\mu = 1$, $B = H$, and in addition to that takes account of the occurrence of the di-electric constant ϵ, and conductivity σ according to equations

(eqXXXIV″, p. 327) $D - E = (\epsilon - 1)\{E + [uB]\}$

(eqXXXIII′, p. 223) $J = \sigma(E + [uB])$

Lorentz's E, D, H are here denoted by E, M, e, m while J denotes the conduction current.

The three last equations whioh have been just cited here coincide with eqⁿ (II), (III), (IV), the first equation would be, if J is identified with $C, = u\rho$ (the current being zero for $\sigma = 0$,

(29) $\text{Curl}\ [H - (u, E)] = C + \frac{dD}{dt} - \text{curl}\ [uD],$

and thus comes out to be in a different form than (1) here. Therefore for magnetised bodies, Lorentz's equations do not correspond to the Relativity Principle.

On the other hand, the form corresponding to the relativity principle, for the condition of non-magnetisation is to be taken out of (D) in §8, with $\mu=1$, not as $B=H$, as Lorentz takes, but as (30) $B-[uD]=H-[uD]$ $(M-[uE]=m-[ue]$ Now by putting $H=B$, the differential equation (29) is transformed into the same form as eqn (1) here when $m-[ue]=M-[uE]$. Therefore it so happens that by a compensation of two contradictions to the relativity principle, the differential equations of Lorentz for moving non-magnetised bodies at last agree with the relativity postulate.

If we make use of (30) for non-magnetic bodies, and put accordingly $H=B+[u, (D-E)]$, then in consequence of (C) in §8,

$$(\epsilon-1)\,(E+[u,B])=D-E+[u.\,[u,D-E]],$$

i.e. for the direction of u

$$(\epsilon-1)\,(E+[uB])_u=(D-E)_u$$

and for a perpendicular direction ū,

$$(\epsilon-1)\,[E+(uB)]_{\bar{u}}=(1-u^2)\,(D-E)_{\bar{u}}$$

i.e. it coincides with Lorentz's assumption, if we neglect u^2 in comparison to 1.

Also to the same order of approximation, Lorentz's form for J corresponds to the conditions imposed by the relativity principle [comp. (E) § 8]—that the components of J_u, $J_{\bar{u}}$ are equal to the components of $\sigma\,(E+[u\,B])$ multiplied by $\sqrt{1-u^2}$ or $\dfrac{1}{\sqrt{1-u^2}}$ respectively.

§10. Fundamental Equations of E. Cohn.

E. Cohn assumes the following fundamental equations.

$$(31)\quad \text{Curl}\,(\text{M}+[v\ \text{E}])=\frac{d\text{E}}{dt}+\text{u div. E}+\text{J}$$

$$-\text{Curl}\ [\text{E}-(v.\ \text{M})]=\frac{d\text{M}}{dt}+\text{u div. M.}$$

$$(32)\quad \text{J}=\sigma\ \text{E},\ =\epsilon\,\text{E}-[v\ \text{M}],\ \text{M}=\mu\,(m+[v\ \text{E.}])$$

where E M are the electric and magnetic field intensities (forces), E, M are the electric and magnetic polarisation (induction). The equations also permit the existence of true magnetism; if we do not take into account this consideration, div. M. is to be put $=o$.

An objection to this system of equations is that according to these, for $\epsilon=1$, $\mu=1$, the vectors force and induction do not coincide. If in the equations, we conceive E and M and not E$-$(U. M), and M$+$[U E] as electric and magnetic forces, and with a glance to this we substitute for E, M, E, M, div. E, the symbols e, M, E $+$[U M], $m-[v\ e]$, ρ, then the differential equations transform to our equations, and the conditions (32) transform into

$$\text{J}=\sigma(\text{E}+[v\ \text{M}])$$

$$e+[u,\ (m-[v\ e])\ ]\ =\ \epsilon(\text{E}+[v\,\text{M}])$$

$$\text{M}-[v,\ (\text{E}+v\ \text{M})]=\mu(m-[u\ e]\ )$$

then in fact the equations of Cohn become the same as those required by the relativity principle, if errors of the order v^2 are neglected in comparison to 1.

It may be mentioned here that the equations of Hertz become the same as those of Cohn, if the auxilliary conditions are

$$(33)\quad \text{E}=\epsilon\,\text{E},\ \text{M}=\mu\text{M},\ \text{J}=\sigma\text{E}.$$

§11. Typical Representations of the Fundamental Equations.

In the statement of the fundamental equations, our leading idea had been that they should retain a covariance of form, when subjected to a group of Lorentz-transformations. Now we have to deal with ponderomotive reactions and energy in the electro-magnetic field. Here from the very first there can be no doubt that the settlement of this question is in some way connected with the simplest forms which can be given to the fundamental equations, satisfying the conditions of covariance. In order to arrive at such forms, I shall first of all put the fundamental equations in a typical form which brings out clearly their covariance in case of a Lorentz-transformation. Here I am using a method of calculation, which enables us to deal in a simple manner with the space-time vectors of the 1st, and 2nd kind, and of which the rules, as far as required are given below.

A system of magnitudes a_{hk} formed into the matrix

$$\left|\begin{matrix} a_{11} & \cdots\cdots\cdots & a_{1q} \\ & & \\ & & \\ a_{p1} & \cdots\cdots\cdots & a_{pq} \end{matrix}\right|$$

arranged in p horizontal rows, and q vertical columns is called a $p \times q$ series-matrix, and will be denoted by the letter A.

If all the quantities a_{hk} are multiplied by C, the resulting matrix will be denoted by CA.

If the roles of the horizontal rows and vertical columns be intercharged, we obtain a $q \times p$ series matrix, which will be known as the transposed matrix of A, and will be denoted by $\bar{A}$.

$$\bar{A} = \begin{vmatrix} a_{11} & \cdots & a_{p1} \\ \vdots & & \vdots \\ a_{1q} & \cdots & a_{pq} \end{vmatrix}$$

If we have a second $p \times q$ series matrix B,

$$B = \begin{vmatrix} b_{11} & \cdots & b_{1q} \\ \vdots & & \vdots \\ b_{p1} & \cdots & b_{pq} \end{vmatrix}$$

then $A + B$ shall denote the $p \times q$ series matrix whose members are $a_{hk} + b_{hk}$.

2° If we have two matrices

$$A = \begin{vmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pq} \end{vmatrix} \qquad B = \begin{vmatrix} b_{11} & \cdots & b_{1r} \\ \vdots & & \vdots \\ b_{q1} & \cdots & b_{pr} \end{vmatrix}$$

where the number of horizontal rows of B, is equal to the number of vertical columns of A, then by AB, the product of the matrics A and B, will be denoted the matrix

$$C = \begin{vmatrix} c_{11} & \cdots & c_{1r} \\ \vdots & & \vdots \\ c_{pr} & \cdots & c_{pp} \end{vmatrix}$$

where $c_{hk} = a_{h1}\, b_{1k} + a_{h2}\, b_{2k} + \ldots\ldots a_{ks}\, b_{sk} + \ldots a_{kq}\, b_{qk}$ these elements being formed by combination of the horizontal rows of A with the vertical columns of B. For such a point, the associative law (AB) S = A(BS) holds, where S is a third matrix which has got as many horizontal rows as B (or AB) has got vertical columns.

For the transposed matrix of C = BA, we have $C = \bar{B}\bar{A}$

3°. We shall have principally to deal with matrices with at most four vertical columns and for horizontal rows.

As a unit matrix (in equations they will be known for the sake of shortness as the matrix 1) will be denoted the following matrix (4 × 4 series) with the elements.

$$(34)\quad \begin{vmatrix} e_{11} & e_{12} & e_{13} & e_{14} \\ e_{21} & e_{22} & e_{23} & e_{24} \\ e_{31} & e_{32} & e_{33} & e_{34} \\ e_{41} & e_{42} & e_{43} & e_{44} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

For a 4 × 4 series-matrix, Det A shall denote the determinant formed of the 4 × 4 elements of the matrix. If det A ‡ o, then corresponding to A there is a reciprocal matrix, which we may denote by A^{-1} so that $A^{-1}A=1$

A matrix

$$f = \begin{vmatrix} o & f_{12} & f_{13} & f_{14} \\ f_{21} & o & f_{23} & f_{24} \\ f_{31} & f_{32} & o & f_{34} \\ f_{41} & f_{42} & f_{43} & o \end{vmatrix}.$$

in which the elements fulfil the relation $f_{hk} = -f_{kh}$, is called an alternating matrix. These relations say that the transposed matrix $\bar{f} = -f$. Then by f^* will be the *dual*, alternating matrix

$$(35)\qquad f^* = \begin{vmatrix} o & f_{34} & f_{42} & f_{23} \\ f_{43} & o & f_{14} & f_{31} \\ f_{24} & f_{41} & o & f_{12} \\ f_{32} & f_{13} & f_{21} & o \end{vmatrix}$$

Then (36) $f^{*} f = f_{34} f_{22} + f_{42} f_{31} + f_{32} f_{14}$

i e. We shall have a 4×4 series matrix in which all the elements except those on the diagonal from left up to right down are zero, and the elements in this diagonal agree with each other, and are each equal to the above mentioned combination in (36).

The determinant of f is therefore the square of the combination, by $\text{Det}^{\frac{1}{2}} f$ we shall denote the expression

$$\text{Det}^{\frac{1}{2}} f = f_{32} f_{14} f_{13} f_{24} + f_{21} f_{34}.$$

4°. A linear transformation

$$x_h = a_{h1} x_1' + a_{h2} x' + a_{h3} x_3' + a_{h4} x_4' \quad (h = 1, 2, 3,$$

which is accomplished by the matrix

$$A = \left| \begin{matrix} a_{11}, & a_{12}, & a_{13}, & a_{14} \\ a_{21}, & a_{\ 2}, & a_{23}, & a_{24} \\ a_{31}, & a_{32}, & a_{33}, & a_{34} \\ a_{41}, & a_{4\,2}, & a_{43}, & a_{44} \end{matrix} \right|$$

wlll be denoted as the transformation A

By the transformation A, the expression $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is changed into the quadratic form $\sum a_{hk} x_h' x_k'$,

where $a_{hk} = a_{1k} a_{1k} + a_{2h} a_{2k} + a_{3h} a_{3k} + a_{4h} a_{4k}$.

are the members of a 4×4 series matrix which is the product of $\bar{A}$ A, the transposed matrix of A into A. If by the transformation, the expression is changed to

$$x_1'^2 + x_2'^2 + x_3'^2 + x_4'^2,$$

we must have $\bar{A}\ A = 1$.

A has to correspond to the following relation, if transformation (38) is to be a Lorentz-transformation. For the determinant of A) it follows out of (39) that $(\text{Det A})^2 = 1$, or Det $A = \pm 1$.

From the condition (39) we obtain

$$A^{-1} = \bar{A},$$

i.e. the reciprocal matrix of A is equivalent to the transposed matrix of A.

For A as Lorentz transformation, we have further Det $A = +1$, the quantities involving the index 4 once in the subscript are purely imaginary, the other co-efficients are real, and $a_{44} > 0$.

5°. A space time vector of the first kind* which s represented by the 1×4 series matrix,

$$(41) \qquad s = | \, s_1 \ \ s_2 \ \ s_3 \ \ s_4 \, |$$

is to be replaced by sA in case of a Lorentz transformation A. *i.e.* $s' = | \, s_1' \ \ s_2' \ \ s_3' \ \ s_4' \, | = | \, s_1 \ \ s_2 \ \ s_3 \ \ s_4 \, | \, A$; A space-time vector of the 2nd kind† with components $f_{23} \cdots f_{34}$ shall be represented by the alternating matrix

$$(42) \qquad f = \begin{vmatrix} o & f_{12} & f_{13} & f_{14} \\ f_{21} & o & f_{23} & f_{24} \\ f_{31} & f_{32} & o & f_{34} \\ f_{41} & f_{42} & f_{43} & o \end{vmatrix}.$$

and is to be replaced by $A^{-1} f A$ in case of a Lorentz transformation [see the rules in § 5 (23) (24)]. Therefore referring to the expression (37), we have the identity $\text{Det}^{\frac{1}{2}} (\bar{A} f A) = \text{Det A}. \ \text{Det}^{\frac{1}{2}} f$. Therefore $\text{Det}^{\frac{1}{2}} f$ becomes an invariant in the case of a Lorentz transformation [see eq. (26) Sec. § 5].

---

* *Vide* note **13.**
† *Vide* note **14.**

Looking back to (36), we have for the dual matrix
$(\bar{A} f^* A)(A^{-1} f A) = A^{-1} f^* f A = \mathrm{Det}^{\frac{1}{2}} f.\ A^{-1} A = \mathrm{Det}^{\frac{1}{2}} f$
from which it is to be seen that the dual matrix f^* behaves exactly like the primary matrix f, and is therefore a space time vector of the II kind; f^* is therefore known as the dual space-time vector of f with components (f_{14}, f_{24}, f_{34}), (f_{23}, f_{31}, f_{12}).

6.* If w and s are two space-time rectors of the 1st kind then by $w \bar{s}$ (as well as by $s \bar{w}$) will be understood the combination (43) $w_1 s_1 + w_2 s_2 + w_3 s_3 + w_4 s_4$.

In case of a Lorentz transformation A, since $(wA)(\bar{A} \bar{s}) = w \bar{s}$, this expression is invariant.—If $w \bar{s} = 0$, then w and s are perpendicular to each other.

Two space-time rectors of the first kind (w, s) gives us a 2×4 series matrix

$$\left|\begin{array}{cccc} w_1 & w_2 & w_3 & w_4 \\ s_1 & s_2 & s_3 & s_4 \end{array}\right|$$

Then it follows immediately that the systēm of six magnitudes (44) $w_2 s_3 - w_3 s_2,\ w_3 s_1 - w_1 s_3,\ w_1 s_2 - w_2 s_1,$

$w_1 s_4 - w_4 s_1,\ w_2 s_4 - w_4 s_2,\ w_3 s_4 - w_4 s_3,$

behaves in case of a Lorentz-transformation as a space-time vectör of the II kind. The vector of the second kind with the components (44) are denoted by $[w, s]$. We see easily that $\mathrm{Det}^{\frac{1}{2}} [w, s] = 0$. The dual vector of $[w, s]$ shall be written as $[w, s]^*$.

If $\bar{w}$ is a space-time vector of the 1st kind, f of the second kind, wf signifies a 1×4 series matrix. In case of a Lorentz-transformation A, w is changed into $w' = wA$, f into $f' = A^{-1} f A$,—therefore $w' f'$ becomes $= (wA\ A^{-1} f\ A) = wf\,A$ *i.e.* wf is transformed as a space-time vector of the 1st kind.* We can verify, when w is a space-time vector of the 1st kind, f of the 2nd kind, the important identity

(45) $[w, wf] + [w, wf^*]^* = (w\bar{w})f.$

The sum of the two space time vectors of the second kind on the left side is to be understood in the sense of the addition of two alternating matrices.

For example, for $\omega_1 = o$, $\omega_2 = o$, $\omega_3 = o$, $\omega_4 = i$,

$$\omega f = | if_{41}, if_{42}, if_{43}, o| \; ; \; \omega f^* = | if_{32}, if_{13}, if_{21}, o|$$

$$[\omega \cdot \omega f] = o, o, o, f_{41}, f_{42}, f_{43} \; ; \; [\omega \cdot \omega f^*]^* = o, o, o, f_{32}, f_{13}, f_{21}.$$

The fact that in this special case, the relation is satisfied, suffices to establish the theorem (45) generally, for this relation has a covariant character in case of a Lorentz transformation, and is homogeneous in $(\omega_1, \omega_2, \omega_3, \omega_4)$.

After these preparatory works let us engage ourselves with the equations (C,) (D,) (E) by means which the constants ϵ μ, σ will be introduced.

Instead of the space vector u, the velocity of matter, we shall introduce the space-time vector of the first kind ω with the components.

$$\omega_1 = \frac{u_x}{\sqrt{1-u_2}}, \; \omega_2 = \frac{u_y}{\sqrt{1-u^2}}, \; \omega_3 = \frac{u_z}{\sqrt{1-u^2}}; \; \omega_4 = \frac{i}{\sqrt{1-u^2}}.$$

(40) where $\omega_1{}^2 + \omega_2{}^2 + \omega_3{}^2 + \omega_4{}^2 = -1$
and $-i\omega_4 > 0$.

By F and f shall be understood the space time vectors of the second kind $M - iE$, $m - ie$.

In $\Phi = \omega F$, we have a space time vector of the first kind with components

$$\left.\begin{aligned} \Phi_1 &= \qquad\quad \omega_2 F_{12} + \omega_3 F_{13} + \omega_4 F_{14} \\ \Phi_1 &= \omega_1 F_{21} \qquad\quad + \omega_3 F_{23} + \omega_4 F_{24} \\ \Phi_3 &= \omega_1 F_{31} + \omega_2 F_{32} \qquad\quad + \omega_4 F_{34} \\ \Phi_4 &= \omega_1 F_{41} + \omega_2 F_{42} + \omega_3 F_{43} \end{aligned}\right\}$$

* *Vide* note 15.

The first three quantities (ϕ_1, ϕ_2, ϕ_3) are the components of the space-vector $\dfrac{E+[u, M]}{\sqrt{1-u^2}}$.

and further $\phi_4 = \dfrac{i\,[u\,E]}{\sqrt{1-u^2}}$.

Because F is an alternating matrix,

(49) $\omega\bar{\Phi} = \omega_1\phi_1 + \omega_2\Phi_2 + \omega_3\Phi_3 + \omega_4\Phi_4 = o.$

i.e. Φ is perpendicular to the vector ω; we can also write $\Phi_4 = i\,[\omega_x\Phi_1 + \omega_y\Phi_2 + \omega_z\Phi_3]$.

I shall call the space-time vector Φ of the first kind as the *Electric Rest Force.**

Relations analogous to those holding between $-\omega F$, E, M, U, hold amongst $-\omega f$, e, m, u, and in particular $-\omega f$ is normal to ω. The relation (C) can be written as

$$\{C\} \qquad \omega f = \epsilon\omega F.$$

The expression (ωf) gives four components, but the fourth can be derived from the first three.

Let us now form the time-space vector 1st kind $\Psi = i\omega f^*$, whose components are

$$\left.\begin{aligned}
\Psi_1 &= -i\,(\qquad\;\; \omega_2 f_{34} + \omega_3 f_{42} + \omega_4 f_{23}) \\
\Psi_2 &= -i\,(\omega_1 f_{43} + \qquad\;\; \omega_3 f_{14} + \omega_4 f_{31}) \\
\Psi_3 &= -i\,(\omega_1 f_{24} + \omega_2 f_{41} \qquad\;\; + \omega_4 f_{12}) \\
\Psi_4 &= -i\,(\omega_1 f_{32} + \omega_2 f_{13} + \omega_3 f_{21} \qquad\;\;)
\end{aligned}\right\}$$

Of these. the first three Ψ_1, Ψ_2, Ψ_3, are the x, y, z components of the space-vector (51) $\dfrac{m-(uc)}{\sqrt{1-u^2}}$

and further (52) $\psi_4 = \dfrac{i\,(um)}{\sqrt{1-u^2}}$.

---

* *Vide* note 16.

Among these there is the relation

(53) $\bar{\omega\Psi}=\omega_1\Psi_1+\omega_2\Psi_2+\omega_3\Psi_3+\omega_4\Psi_4=o$

which can also be written as $\Psi_4=i\ (u_x\Psi_1+u_y\Psi_2+u_z\Psi_3)$.

The vector Ψ is perpendicular to ω; we can call it the *Magnetic rest-force.*

Relations analogous to these hold among the quantities $\iota\omega F^*$, M, E, u and Relation (D) can be replaced by the formula

$$\{\,D\,\} \qquad -\omega F^*=\mu\omega f^*.$$

We can use the relations (C) and (D) to calculate F and f from Φ and Ψ we have

$$\omega F=-\Phi,\ \omega F^*=-i\mu\Psi,\ \omega f=-\epsilon\Phi,\ \omega f^*=-i\Psi.$$

and applying the relation (45) and (46), we have

$$F=\ [\omega.\ \Phi]+i\mu[\omega.\ \Psi]^* \qquad 55)$$

$$f=\ \epsilon[\omega.\ \Phi]+i[\omega.\ \Psi]^* \qquad 56).$$

i.e. $F_{12}=(\omega_1\Phi_1-\omega_2\Phi_1)+i\mu[\omega_3\Psi_4-\omega_4\Psi_3]$, etc.

$f_{12}=\epsilon(\omega_1\Phi_2-\omega_2\Phi_1)+i\ [\omega_3\Psi_4-\omega_4\Psi_3]$. etc.

Let us now consider the space-time vector of the second kind $[\Phi\ \Psi]$, with the components

$$\begin{bmatrix} \Phi_2\Psi_3-\Phi_3\Psi_2, & \Phi_3\Psi_1-\Phi_1\Psi_3, & \Phi_1\Psi_2-\Phi_2\Psi_1 \\ \Phi_1\Psi_4-\Phi_4\Psi_1, & \Phi_2\Psi_4-\Phi_4\Psi_2; & \Phi_3\Psi_4-\Phi_4\Psi_3 \end{bmatrix}$$

Then the corresponding space-time vector of the first kind $\omega[\Phi,\ \Psi]$ vanishes identically owing to equations 9) and 53)

for $\omega[\Phi.\Psi]=-(\bar{\omega\Psi})\Phi+(\bar{\omega\Phi})\Psi$

Let us now take the vector of the 1st kind

(57 $$\Omega=i\omega[\Phi\Psi]^*$$

with the components $\Omega_1=-i\begin{vmatrix} \omega_2 & \omega_3 & \omega_4 \\ \Phi_2 & \Phi_3 & \Phi_4 \\ \Psi_2 & \Psi_3 & \Psi_4 \end{vmatrix}$. etc.

Then by applying rule (45), we have

(58) $[\Phi.\Psi] = i\,[\omega\Omega]^*$

$i.e.$ $\Phi_1\Psi_2 - \Phi_2\Psi_1 = i(\omega_3\Omega_4 - \omega_4\Omega_3)$ etc.

The vector Ω fulfils the relation

$$(\omega\Omega) = \omega_1\Omega_1 + \omega_2\Omega_2 + \omega_3\Omega_3 + \omega_4\Omega_4 = o.$$

(which we can write as $\Omega_4 = i\,(\omega_x\Omega_1 + \omega_y\Omega_2 + \omega_2\Omega_3)$

and Ω is also normal to ω. In case $\omega = o$,

we have $\Phi_4 = o$, $\Psi_4 = o$, $\Omega_4 = o$, and

$$[\Omega_1, \Omega_2, \Omega_3 = \begin{vmatrix} \Phi_1 & \Phi_2 & \Phi_3 \\ \Psi_1 & \Psi_2 & \Psi_3 \end{vmatrix} .$$

I shall call Ω, which is a space-time vector 1st kind the Rest-Ray.

As for the relation E), which introduces the conductivity σ we have $-\omega S = -(\omega_1 s_1 + \omega_2 s_2 + \omega_3 s_3 + \omega_4 s_4)$

$$= \frac{-\,|\,u\,|\,C_u + \rho}{\sqrt{1-u^2}} = \rho'.$$

This expression gives us the rest-density of electricity (see §8 and §4).

Then $61) = s + (\omega\bar{s})\omega$

represents a space-time vector of the 1st kind, which since $\omega\omega = -1$, is normal to ω, and which I may call the rest-current. Let us now conceive of the first three component of this vector as the $(x - y - z)$ co-ordinates of the space-vector, then the component in the direction of u is

$$C_u - \frac{|\,u\,|\,\rho'}{\sqrt{1-u^2}} = \frac{c_u - |\,u\,|\,\rho}{\sqrt{1-u^2}} = \frac{J_u}{1-u^2} .$$

and the component in a perpendicular direction is $C_{\bar{u}} = J_{\bar{u}}$.

This space-vector is connected with the space-vector $J = C - \rho u$, which we denoted in §8 as the conduction-current.

Now by comparing with $\Phi = -\omega F$, the relation (E) can be brought into the form

$$\{E\} \qquad s + (\omega \bar{s})\omega = -\sigma \omega F,$$

This formula contains four equations, of which the fourth follows from the first three, since this is a space-time vector which is perpendicular to ω.

Lastly, we shall transform the differential equations (A) and (B) into a typical form.

§12. The Differential Operator Lor.

A 4×4 series matrix 62) $S = \begin{vmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{vmatrix} = | S_{kh} |$

with the condition that in case of a Lorentz transformation it is to be replaced by $\bar{A}SA$, may be called a space-time matrix of the II kind. We have examples of this in :—

1) the alternating matrix f, which corresponds to the space-time vector of the II kind,—

2) the product f F of two such matrices, for by a transformation A, it is replaced by $(A^{-1} f A \cdot A^{-1} F A) = A^{-1} f F A$,

3) further when $(\omega_1, \omega_2, \omega_3, \omega_4)$ and $(\Omega_1, \Omega_2, \Omega_3, \Omega_4)$ are two space-time vectors of the 1st kind, the 4×4 matrix with the element $S_{hk} = \omega_h \Omega_k$,

lastly in a multiple L of the unit matrix of 4×4 series in which all the elements in the principal diagonal are equal to L, and the rest are zero.

We shall have to do constantly with functions of the space-time point (x, y, z, it), and we may with advantage employ the 1×4 series matrix, formed of differential symbols,—

$$\left| \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\iota\partial t}, \right| \text{ or } (63) \left| \frac{\partial}{\partial x_1} \; \frac{\partial}{\partial x_2} \; \frac{\partial}{\partial x_3} \; \frac{\partial}{\partial x_4} \right|$$

For this matrix I shall use the shortened from " lor."*

Then if S is, as in (62), a space-time matrix of the II kind, by lor S′ will be understood the 1×4 series matrix

$$| K_1 \quad K_2 \quad K_3 \quad K_4 |$$

where $K_k = \frac{\partial S_{1k}}{\partial x_1} + \frac{\partial S_{2k}}{\partial x_2} + \frac{\partial S_{3k}}{\partial x_3} + \frac{\partial S_{4k}}{\partial x_4}$,

When by a Lorentz transformation A, a new reference system ($x'_1 \; x'_2 \; x'_3 \; x_4$) is introduced, we can use the operator

$$\text{lor}' = \left| \frac{\partial}{\partial x_1'}, \; \frac{\partial}{\partial x_2'}, \; \frac{\partial}{\partial x_3'}, \; \frac{\partial}{\partial x_4'} \right|$$

Then S is transformed to $S' = \bar{A}\, S\, A = |\, S'_{kk} \,|$, so by lor ′S′ is meant the 1×4 series matrix, whose element are

$$K'_k = \frac{\partial S'_{1k}}{\partial x_1'} + \frac{\partial S'_{2k}}{\partial x_2'} + \frac{\partial S'_{3k}}{\partial x_3'} + \frac{\partial S'_{4k}}{\partial x_4'}.$$

Now for the differentiation of any function of ($x \; y \; z \; t$) we have the rule

$$\frac{\partial}{\partial x_k'} = \frac{\partial}{\partial x_1} \frac{\partial x_1}{\partial x_k'} + \frac{\partial}{\partial x_2} \frac{\partial x_2}{\partial x_k'}$$

$$+ \frac{\partial}{\partial x_3} \frac{\partial x_3}{\partial x_k'} + \frac{\partial}{\partial x_4} \frac{\partial x_4}{\partial x'_k}$$

$$= \frac{\partial}{\partial x_1} a_{1k} + \frac{\partial}{\partial x_2} a_{2k} + \frac{\partial}{\partial x_3} a_{3k} + \frac{\partial}{\partial x_4} a_{4k}.$$

so that, we have symbolically lor′=lor A.

---

* *Vide* note 17.

Therefore it follows that

$$\text{lor}\ 'S' = \text{lor}\ (A\ A^{-1}\ SA) = (\text{lor}\ S)A.$$

i.e., lor S behaves like a space-time vector of the first kind.

If L is a multiple of the unit matrix, then by lor L will be denoted the matrix with the elements

$$\left| \frac{\partial L}{\partial x_1} \quad \frac{\partial L}{\partial x_2} \quad \frac{\partial L}{\partial x_3} \quad \frac{\partial L}{\partial x_4} \right|$$

If s is a space-time vector of the 1st kind, then

$$\text{lor}\ \bar{s} = \frac{\partial s_1}{\partial x_1} + \frac{\partial s_2}{\partial x_2} + \frac{\partial s_3}{\partial x_3} + \frac{\partial s_4}{\partial x_4}.$$

In case of a Lorentz transformation A, we have

$$\text{lor}\ '\bar{s}' = \text{lor}\ A.\quad \bar{A}s = \text{lor}\ s.$$

i.e., lor s is an invariant in a Lorentz-transformation.

In all these operations the operator lor plays the part of a space-time vector of the first kind.

If f represents a space-time vector of the second kind, $-$lor f denotes a space-time vector of the first kind with the components

$$\begin{array}{llll}
 & \dfrac{\partial f_{12}}{\partial x_2} & +\dfrac{\partial f_{13}}{\partial x_3} & +\dfrac{\partial f_{14}}{\partial x_4}, \\[2ex]
\dfrac{\partial f_{21}}{\partial x_1} & & +\dfrac{\partial f_{23}}{\partial x_3} & +\dfrac{\partial f_{24}}{\partial x_4}, \\[2ex]
\dfrac{\partial f_{31}}{\partial x_1} & +\dfrac{\partial f_{32}}{\partial x_2} & & +\dfrac{\partial f_{34}}{\partial x_4}, \\[2ex]
\dfrac{\partial f_{41}}{\partial x_1} & +\dfrac{\partial f_{42}}{\partial x_2} & +\dfrac{\partial f_{43}}{\partial x_3} &
\end{array}$$

So the system of differential equations (A) can be expressed in the concise form

$$\{A\} \qquad \text{lor}\, f = -s,$$

and the system (B) can be expressed in the form

$$\{B\} \qquad \log F^* = 0.$$

Referring back to the definition (67) for log $\bar{s}$, we find that the combinations lor $(\overline{\text{lor}\, f})$, and lor $(\overline{\text{lor}\, F^*}$ vanish identically, when f and F^* are alternating matrices. Accordingly it follows out of $\{A\}$, that

$$(68) \qquad \frac{\partial s_1}{\partial x_1} + \frac{\partial s_2}{\partial x_2} + \frac{\partial s_3}{\partial x_3} + \frac{\partial s_4}{\partial x_4} = 0,$$

while the relation

(69) lor (lor F^*) = 0, signifies that of the four equations in $\{B\}$, only three represent independent conditions.

I shall now collect the results.

Let ω denote the space-time vector of the first kind

$$\left(\frac{u}{\sqrt{1-u^2}}, \frac{i}{\sqrt{1-u^2}}\right)$$

(u = velocity of matter),

F the space-time vector of the second kind (M, $-i$E)

(M = magnetic induction, E = Electric force,

f the space-time vector af the second kind (m, $-ie$)

(m = magnetic force, e = Electric Induction.

s the space-time vector of the first kind (C, $i\rho$)

(ρ = electrical space-density, C $- \rho u$ = conductivity current,

ϵ = dielectric constant, μ = magnetic permeability,

σ = conductivity,

then the fundamental equations for electromagnetic processes in moving bodies are*

$\{A\}$ $\mathrm{lor}\, f = -s$

$\{B\}$ $\log F^* = o$

$\{C\}$ $\omega f = \epsilon\omega F$

$\{D\}$ $\omega F^* = \mu\omega f^*$

$\{E\}$ $s + (\omega \bar{s}),\ w = -\sigma\omega F.$

$\omega\bar{\omega} = -1$, and ωF, ωf, ωF^*, ωf^*, $s+(\omega\bar{s})\omega$ which are space-time vectors of the first kind are all normal to ω, and for the system $\{B\}$, we have

$$\mathrm{lor}\ (\mathrm{lor}\ F^*) = 0.$$

Bearing in mind this last relation, we see that we have as many independent equations at our disposal as are necessary for determining the motion of matter as well as the vector u as a function of x, y, z, t, when proper fundamental data are given.

§ 13. The Product of the Field-vectors fF.

Finally let us enquire about the laws which lead to the determination of the vector ω as a function of $(x,y,z,t.)$ In these investigations, the expressions which are obtained by the multiplication of two alternating matrices

$$f = \begin{vmatrix} 0 & f_{12} & f_{13} & f_{14} \\ f_{21} & 0 & f_{23} & f_{24} \\ f_{31} & f_{32} & 0 & f_{34} \\ f_{41} & f_{42} & f_{43} & 0 \end{vmatrix} \qquad F = \begin{vmatrix} 0 & F_{12} & F_{13} & F_{14} \\ F_{21} & 0 & F_{23} & F_{24} \\ F_{31} & F_{32} & 0 & F_{34} \\ F_{41} & F_{42} & F_{43} & 0 \end{vmatrix}$$

* *Vide* note 19.

are of much importance. Let us write.

$$(70)\quad fF=\begin{vmatrix} S_{11}-L & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22}-L & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33}-L & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44}-L \end{vmatrix}$$

Then (71) $S_{11}+S_{22}+S_{33}+S_{44}=0.$

Let L now denote the symmetrical combination of the indices 1, 2, 3, 4, given by

$$(72)\quad L=\frac{1}{2}\Big(f_{23}F_{23}+f_{31}F_{31}+f_{12}F_{12}+f_{14}F_{14}+f_{24}F_{24}+f_{34}F_{34}\Big)$$

Then we shall have

$$(73)\quad S_{11}=\frac{1}{2}\Big(f_{23}F_{23}+f_{34}F_{34}+f_{42}F_{42}-f_{12}F_{12}-f_{13}F_{13}\quad f_{14}F_{14}\Big)$$

$$S_{12}=f_{13}F_{32}+f_{14}F_{42}\ \text{etc.} \ldots$$

In order to express in a real form, we write

$$(74)\ S=\begin{vmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{vmatrix}=\begin{vmatrix} X_x & Y_x & Z_x & -iT_x \\ X_y & Y_y & Z_y & -iT_y \\ X_z & Y_z & Z_z & -iT_z \\ -iX_t & -iY_t & -iZ_t & T_t \end{vmatrix}$$

Now $X_x=\frac{1}{2}\Big[m_xM_x-m_yM_y-m_zM_z+e_xE_x-e_yE_y-e_zE_z\Big]$

$$(75)\quad X_y = m_x M_y + e_y E_x, \quad Y_x = m_y M_x + e_x E_y \text{ etc.} \qquad *$$

$$X_t = e_y M_z - e_z M_y, \quad T_x = m_x E_y - m_y E_z \text{ etc.}$$

$$T_t = \frac{1}{2}\left[m_x M_x + m_y M_y + m_z M_z + e_x E_x + e_y E_y + e_z E_z\right]$$

$$L_t = \frac{1}{2}\left[m_x M_x + m_y M_y + m_z M_z - e_x E_x - e_y E_y - e_z E_z\right]$$

These quantities are all real. In the theory for bodies at rest, the combinations (X_x, X_y, X_z, Y_x, Y_y, Y_z, Z_x, Z_y, Z_z) are known as "Maxwell's Stresses," T_x, T_y, T_z are known as the Poynting's Vector, T_t as the electromagnetic energy-density, and L as the Langrangian function.

On the other hand, by multiplying the alternating matrices of f^* and F^*, we obtain

$$(77)\quad F^*f^* = \begin{vmatrix} -S_{11}-L, & -S_{12}, & -S_{13}, & -S_{14} \\ -S_{21}, & -S_{22}-L, & -S_{23}, & -S_{24} \\ -S_{31} & -S_{32}, & -S_{33}-L, & -S_{34} \\ -S_{41} & -S_{42} & -S_{43} & -S_{44}-L \end{vmatrix}$$

and hence, we can put

$$(78)\qquad fF = S - L, \qquad F^*f^* = -S - L,$$

where by L, we mean L-times the unit matrix, *i.e.* the matrix with elements

$$|\, Le_{hk} \,|, \quad (e_{hh} = 1, \quad e_{hk} = 0, \quad h \neq k \quad h, k = 1, 2, 3, 4).$$

Since here $SL = LS$, we deduce that,

$$F^* f^* f F = (-S-L)(S-L) = -SS + L^2,$$

and find, since $f^* f = \text{Det}^{\frac{1}{2}} f$, $F^* F = \text{Det}^{\frac{1}{2}} F$, we arrive at the interesting conclusion

* *Vide* note 18.

(79) $$SS = L^2 - \mathrm{Det}^{\frac{1}{2}} f\, \mathrm{Det}^{\frac{1}{2}} F$$

i.e. the product of the matrix S into itself can be expressed as the multiple of a unit matrix—a matrix in which all the elements except those in the principal diagonal are zero, the elements in the principal diagonal are all equal and have the value given on the right-hand side of (79). Therefore the general relations

(80) $$S_{h1}\, S_{1k} + S_{h2}\, S_{2k} + S_{h3}\, S_{3k} + S_{h4}\, S_{4k} = o,$$

h, k being unequal indices in the series 1, 2, 3, 4, and

(81) $$S_{h1}\, S_{1h} + S_{h2}\, S_{2h} + S_{h3}\, S_{3h} + S_{h4}\, S_{4h} = L^2 - \mathrm{Det}^{\frac{1}{2}} f\, \mathrm{Det}^{\frac{1}{2}} F,$$

for $h = 1, 2, 3, 4$.

Now if instead of F, and f in the combinations (72) and (73), we introduce the electrical rest-force Φ, the magnetic rest-force Ψ, and the rest-ray Ω [(55), (56) and (57)], we can pass over to the expressions,—

(82) $$L = -\tfrac{1}{2}\epsilon\, \Phi\overline{\Phi} + \tfrac{1}{2}\mu\, \Psi\overline{\Psi},$$

(83) $$S_{hk} = -\tfrac{1}{2}\epsilon\, \Phi\overline{\Phi}\, e_{hk} - \tfrac{1}{2}\mu\, \Psi\overline{\Psi}\, e_{hk}$$
$$+ \epsilon\, (\Phi_h\, \Phi_k - \Phi\overline{\Phi}\, \omega_h\, \omega_k)$$
$$+ \mu\, (\Psi_h\, \Psi_k - \Psi\overline{\Psi}\, \omega_h\, \omega_k) - \Omega_h\, \omega_k - \epsilon\mu\, \omega_h\, \Omega_k$$
$$(h_1\ k = 1, 2, 3, 4).$$

Here we have

$$\Phi\overline{\Phi} = \Phi_1{}^2 + \Phi_2{}^2 + \Phi_3{}^2 + \Phi_4{}^2,\ \Psi\overline{\Psi} = \Psi_1{}^2 + \Psi_2{}^2 + \Psi_3{}^2 + \Psi_4{}^2$$
$$e_{hh} = 1,\ e_{hk} = o\ (h \neq k).$$

The right side of (82) as well as L is an invariant in a Lorentz transformation, and the 4×4 element on the right side of (83) as well as S_{kh} represent a space time vector of the second kind. Remembering this fact, it suffices, for establishing the theorems (82) and (83) generally, to prove it for the special case $\omega_1 = o$, $\omega_2 = o$, $w_3 = o$, $\omega_4 = i$. But for this case $\omega = o$, we immediately arrive at the equations (82) and (83) by means (45), (51), (60) on the one hand, and $e = \epsilon E$, $M = \mu m$ on the other hand.

The expression on the right-hand side of (81), which equals

$$[\tfrac{1}{2}(m\,M - eE)^2] + (em)\,(EM),$$

is $\geqq o$, because $(em = \epsilon\,\Phi\,\bar{\Psi}$, $(EM) = \mu\,\Phi\,\bar{\Psi}$; now referring back to 79), we can denote the positive square root of this expression as $\text{Det}^{\frac{1}{4}}\,S$.

Since $\bar{f} = -f$, and $\bar{F} = - F$, we obtain for $\bar{S}$, the transposed matrix of S, the following relations from (78),

$$(84) \qquad F f = \bar{S} - L,\; f^*\,F^* = -\bar{S} - L,$$

Then is $\bar{S} - S = |\, S_{hk} - S_{lk} \,|$

an alternating matrix, and denotes a space-time vector of the second kind. From the expressions (83), we obtain,

$$(85) \qquad S - \bar{S} = -(\epsilon\,\mu - 1)\,[\omega, \Omega],$$

from which we deduce that [see (57), (58)].

$$(86) \qquad \omega\,(S - \bar{S})^* = o,$$

$$(87) \qquad \omega\,(S - \bar{S}) = (\epsilon\,\mu - 1)\,\Omega$$

When the matter is at rest at a space-time point, $\omega = o$, then the equation 86) denotes the existence of the following equations

$$Z_y = Y_z, \quad X_z = Z_x, \quad Y_x = X_y,$$

and from 83),

$$T_x=\Omega_1,\ T_y=\Omega_2,\ T_z=\Omega_3$$

$$X_t=\epsilon\mu\Omega_1,\ Y_t=\epsilon\mu\Omega_2,\ Z_t=\epsilon\mu\Omega_3$$

Now by means of a rotation of the space co-ordinate system round the null-point, we can make,

$$Z_y=Y_z=o,\quad X_z=Z_x=o,\quad X_x=X_y=o.$$

According to 71), we have

$$(88)\quad X_x+Y_y+Z_z+T_t=o,$$

and according to 83), $T_t>o$. In special cases, where Ω vanishes it follows from 81) that

$$X_x{}^2=Y_y{}^2=Z_z{}^2=T_t{}^2,=(\mathrm{Det}^{\frac{1}{4}}\,S)^2,$$

and if T_t and one of the three magnitudes X_x, Y_y, Z_z are $=\pm$ Det$^{\frac{1}{4}}$ S, the two others$=-$ Det$^{\frac{1}{4}}$ S. If Ω does not vanish let $\Omega\neq 0$, then we have in particular from 80)

$$T_z\,X_t=0,\ T_z\,Y_t=0,\ Z_zT_z+T_zT_t=0,$$

and if $\Omega_1=0$, $\Omega_2=0$, $Z_z=-T_t$ It follows from (81), (see also 83) that

$$X_x=-Y_y=\pm\mathrm{Det}^{\frac{1}{4}}\,S,$$

and $$-Z_z=T_t=\sqrt{\mathrm{Det}^{\frac{1}{2}}\,S+\epsilon\mu\Omega_3{}^2}\quad>\mathrm{Det}^{\frac{1}{4}}S.—$$

The space-time vector of the first kind

$$(89)\qquad K=\mathrm{lor}\ S,$$

is of very great importance for which we now want to demonstrate a very important transformation

According to 78), $S=L+fF$, and it follows that

$$\mathrm{lor}\ S=\mathrm{lor}\ L+\mathrm{lor}\ fF.$$

The symbol 'lor' denotes a differential process which in lor fF, operates on the one hand upon the components of f, on the other hand also upon the components of F. Accordingly lor fF can be expressed as the sum of two parts. The first part is the product of the matrices (lor f) F, lor f being regarded as a 1×4 series matrix. The second part is that part of lor fF, in which the diffentiations operate upon the components of F alone. From 78) we obtain

$$f\,\mathrm{F} = -\mathrm{F}^* f^* - 2\mathrm{L};$$

hence the second part of lor fF $= -$(lor F*)f^*+ the part of -2 lor L, in which the differentiations operate upon the components of F alone. We thus obtain

$$\text{lor } \mathrm{S} = (\text{lor } f)\mathrm{F} - (\text{lor } \mathrm{F}^*) f^* + \mathrm{N},$$

where N is the vector with the components

$$\mathrm{N}_h = \tfrac{1}{2}\Big(\frac{\partial f_{23}}{\partial x_h}\mathrm{F}_{23} + \frac{\partial f_{31}}{\partial x_h}\mathrm{F}_{31} + \frac{\partial f_{12}}{\partial x_h}\mathrm{F}_{12} + \frac{\partial f_{14}}{\partial x_h}\mathrm{F}_{14} + \frac{\partial f_{24}}{\partial x_h}\mathrm{F}_{21} + \frac{\partial f_{34}}{\partial x_h}\mathrm{F}_{34}$$
$$- \frac{\partial \mathrm{F}_{23}}{\partial x_h} f_{23} - \frac{\partial \mathrm{F}_{31}}{\partial x_h} f_{31} - \frac{\partial \mathrm{F}_{12}}{\partial x_h} f_{12} - \frac{\partial \mathrm{F}_{14}}{\partial x_h} f_{14} - \frac{\partial \mathrm{F}_{24}}{\partial x_h} f_{24} - \frac{\partial \mathrm{F}_{34}}{\partial x_h} f_{3-} \Big).$$

$$(h = 1, 2, 3, 4)$$

By using the fundamental relations A) and B), 90) is transformed into the fundamental relation

$$(91) \qquad \text{lor } \mathrm{S} = -s\mathrm{F} + \mathrm{N}.$$

In the limitting case $\epsilon = 1$, $\mu = 1$, $f = \mathrm{F}$, N vanishes identically.

Now upon the basis of the equations (55) and (56), and referring back to the expression (82) for L, and from 57) we obtain the following expressions as components of N,—

$$(92)\qquad N_h = -\frac{1}{2}\,\Phi\bar{\Phi}\,\frac{\partial\epsilon}{\partial x_h} - \frac{1}{2}\,\Psi\bar{\Psi}\,\frac{\partial\mu}{\partial x_h}$$

$$+(\epsilon\mu-1)\left(\Omega_1\frac{\partial\omega_1}{\partial x_h}+\Omega_2\frac{\partial\omega_2}{\partial x_h}+\Omega_3\frac{\partial\omega_3}{\partial x_h}+\Omega_4\frac{\partial\omega_4}{\partial x_h}\right)$$

for $h=1, 2, 3, 4$.

Now if we make use of (59), and denote the space-vector which has Ω_1, Ω_2, Ω_3 as the x, y, z components by the symbol W, then the third component of 92) can be expressed in the form

$$(93)\qquad \frac{\epsilon\mu-1}{\sqrt{1-u^2}}\left(W\,\frac{\partial u}{\partial x_h}\right),$$

The round bracket denoting the scalar product of the vectors within it.

§ 14. The Ponderomotive Force.*

Let us now write out the relation K=lor S=$-s$F+N in a more practical form; we have the four equations

$$(94)\quad K_1 = \frac{\partial X_x}{\partial x}+\frac{\partial X_y}{\partial y}+\frac{\partial X_z}{\partial z}-\frac{\partial X_t}{\partial t} = \rho E_x + s_y M_z - s_z M_y$$

$$-\frac{1}{2}\,\Phi\bar{\Phi}\,\frac{\partial\epsilon}{\partial x}-\frac{1}{2}\,\Psi\bar{\Psi}\,\frac{\partial\mu}{\partial x}+\frac{\epsilon\mu-1}{\sqrt{1-u^2}}\left(W\frac{\partial u}{\partial x}\right),$$

$$(95)\quad K_2 = \frac{\partial Y_x}{\partial x}+\frac{\partial Y_y}{\partial y}+\frac{\partial Y_z}{\partial z}-\frac{\partial Y_t}{\partial t} = \rho E_y + s_z M_x - s_x M_z$$

$$-\frac{1}{2}\,\Phi\bar{\Phi}\,\frac{\partial\epsilon}{\partial y}-\frac{1}{2}\,\Psi\bar{\Psi}\,\frac{\partial\mu}{\partial y}+\frac{\epsilon\mu-1}{\sqrt{1-u^2}}\left(W\frac{\partial u}{\partial y}\right),$$

* Vide note 40.

$$(96)\quad K_3 = \frac{\partial Z_x}{\partial x} + \frac{\partial Z_y}{\partial y} + \frac{\partial Z_z}{\partial z} - \frac{\partial Z_t}{\partial t} = \rho E_z + s_x M_y - s_y M_4$$

$$-\frac{1}{2}\Phi\bar{\Phi}\frac{\partial \epsilon}{\partial z} - \frac{1}{2}\Psi\bar{\Psi}\frac{\partial \mu}{\partial z} + \frac{\epsilon\mu - 1}{\sqrt{1-u^2}}\left(W\frac{\partial u}{\partial z}\right),$$

$$(97)\quad \frac{1}{i}K_4 = \frac{\partial T_y}{\partial x} - \frac{\partial T_y}{\partial y} - \frac{\partial T_z}{\partial z} - \frac{\partial T_t}{\partial t} = s_x E_x + s_y E_y + s_z E_z$$

$$-\frac{1}{2}\Phi\bar{\Phi}\frac{\partial \epsilon}{\partial t} - \frac{1}{2}\Psi\bar{\Psi}\frac{\partial \mu}{\partial t} + \frac{\epsilon\mu - 1}{\sqrt{1-u^2}}\left(W\frac{\partial u}{\partial t}\right).$$

It is my opinion that when we calculate the ponderomotive force which acts upon a unit volume at the space-time point x, y, z, t, it has got, x, y, z components as the first three components of the space-time vector

$$K + (\omega \bar{K})\omega,$$

This vector is perpendicular to ω; the law of Energy finds its expression in the fourth relation.

The establishment of this opinion is reserved for a separate tract.

In the limitting case $\epsilon = 1$, $\mu = 1$, $\sigma = 0$, the vector $N = 0$, $S = \rho\omega$, $\omega K = 0$, and we obtain the ordinary equations in the theory of electrons.

APPENDIX

Mechanics and the Relativity-Postulate.

It would be very unsatisfactory, if the new way of looking at the time-concept, which permits a Lorentz transformation, were to be confined to a single part of Physics.

Now many authors say that classical mechanics stand in opposition to the relativity postulate, which is taken to be the basis of the new Electro-dynamics.

In order to decide this let us fix our attention upon a special Lorentz transformation represented by (10), (11), (12), with a vector v in any direction and of any magnitude $q<1$ but different from zero. For a moment we shall not suppose any special relation to hold between the unit of length and the unit of time, so that instead of t, t', q, we shall write ct, ct', and q/c, where c represents a certain positive constant, and q is $<c$. The above mentioned equations are transformed into

$$r'_{\bar{v}}=r_{\bar{v}},\ r'_v=\frac{c\,(r_v-qt)}{\sqrt{c^2-q^2}},\quad t'=\frac{qr_v+c^2t}{c\sqrt{c^2-q^2}}$$

They denote, as we remember, that r is the space-vector (x, y, z), r' is the space-vector $(x'\ y'\ z')$

If in these equations, keeping v constant we approach the limit $c=\infty$, then we obtain from these

$$r'_{\bar{v}}=r_{\bar{v}}.\ r'_v=r_v-qt \qquad t'=t.$$

The new equations would now denote the transformation of a spatial co-ordinate system (x, y, z) to another spatial co-ordinate system $(x'\ y'\ z')$ with parallel axes, the null point of the second system moving with constant velocity in a straight line, while the time parameter remains unchanged. We can, therefore, say that classical mechanics postulates a covariance of Physical laws for the group of homogeneous linear transformations of the expression

$$-x^2-y^2-z^2+c^2 \quad \dots \quad \dots \quad (1)$$

when $c=\infty$.

Now it is rather confusing to find that in one branch of Physics, we shall find a covariance of the laws for the transformation of expression (1) with a finite value of c, in another part for $c=\infty$.

It is evident that according to Newtonian Mechanics, this covariance holds for $c=\infty$, and not for c=volocity of light.

May we not then regard those traditional covariances for $c=\infty$ only as an approximation consistent with experience, the actual covariance of natural laws holding for a certain finite value of c.

I may here point out that by if instead of the Newtonian Relativity-Postulate with $c=\infty$, we assume a relativity-postulate with a finite c, then the axiomatic construction of Mechanics appears to gain considerably in perfection.

The ratio of the time unit to the length unit is chosen in a manner so as to make the velocity of light equivalent to unity.

While now I want to introduce geometrical figures in the manifold of the variables (x, y, z, t), it may be convenient to leave (y, z) out of account, and to treat x and t as any possible pair of co-ordinates in a plane, refered to oblique axes.

A space time null point 0 ($x, y, z, t=0, 0, 0, 0$) will be kept fixed in a Lorentz transformation.

The figure $-x^2-y^2-z^2+t^2=1,\ t>0$ ... ... (2)
which represents a hyper boloidal shell, contains the space-time points A ($x, y, z, t=0, 0, 0, 1$), and all points A′ which after a Lorentz-transformation enter into the newly introduced system of reference as ($x', y', z', t'=0, 0, 0, 1$).

The direction of a radius vector 0A′ drawn from 0 to the point A′ of (2), and the directions of the tangents to (2) at A′ are to be called normal to each other.

Let us now follow a definite position of matter in its course through all time t. The totality of the space-time points (x, y, z, t) which correspond to the positions at different times t, shall be called a space-time line.

The task of determining the motion of matter is comprised in the following problem:—It is required to establish for every space-time point the direction of the space-time line passing through it.

To transform a space-time point P (x, y, z, t) to rest is equivalent to introducing, by means of a Lorentz transformation, a new system of reference (x', y', z', t'), in which the t' axis has the direction 0A′, 0A′ indicating the direction of the space-time line passing through P. The space t'=const, which is to be laid through P, is the one which is perpendicular to the space-time line through P.

To the increment dt of the time of P corresponds the increment

$$d\tau=\sqrt{dt^2-dx^2-dy^2-dz^2}=dt\sqrt{1-u^2}$$

of the newly introduced time parameter t'. The value of the integral

$$\int d\tau=\int\sqrt{-(dx_1^2+dx_2^2+dx_3^2+dx_4^2)}$$

when calculated upon the space-time line from a fixed initial point P_0 to the variable point P, (both being on the space-time line), is known as the 'Proper-time' of the position of matter we are concerned with at the space-time point P. (It is a generalization of the idea of Positional-time which was introduced by Lorentz for uniform motion.)

If we take a body R° which has got extension in space at time t_0, then the region comprising all the space-time line passing through R° and t_0 shall be called a space-time filament.

If we have an anatylical expression $\theta(x\ y, z, t)$ so that $\theta(x, y\ z\ t)=0$ is intersected by every space time line of the filament at one point,—whereby

$$-\left(\frac{\partial\Theta}{\partial x}\right)^2, -\left(\frac{\partial\Theta}{\partial y}\right), -\left(\frac{\partial\Theta}{\partial z}\right)^2, -\left(\frac{\partial\Theta}{\partial t}\right)^2>0, \frac{\partial\Theta}{\partial t}>0.$$

then the tolality of the intersecting points will be called a cross section of the filament.

At any point P of such across-section, we can introduce by means of a Lorentz transformation a system of reference $(x', y, z'\ t)$, so that according to this

$$\frac{\partial\Theta}{\partial x'}=0, \frac{\partial\Theta}{\partial y'}=0, \frac{\partial\Theta}{\partial z'}=0, \frac{\partial\Theta}{\partial t'}>0.$$

The direction of the uniquely determined t'—axis in question here is known as the upper normal of the cross-section at the point P and the value of $dJ=\int\int\int dx'\ dy'\ dz'$ for the surrounding points of P on the cross-section is known as the elementary contents (Inhalts-element) of the cross-section. In this sense R° is to be regarded as the cross-section normal to the t axis of the filament at the point $t=t^{\circ}$, and the volume of the body R° is to be regarded as the contents of the cross-section.

If we allow R^{o} to converge to a point, we come to the conception of an infinitely thin space-time filament. In such a case, a space-time line will be thought of as a principal line and by the term 'Proper-time' of the filament will be understood the 'Proper-time' which is laid along this principal line; under the term normal cross-section of the filament, we shall understand the cross-section upon the space which is normal to the principal line through P.

We shall now formulate the principle of conservation of mass.

To every space R at a time t, belongs a positive quantity—the mass at R at the time t. If R converges to a point (x, y, z, t), then the quotient of this mass, and the volume of R approaches a limit $\mu(x, y, z, t)$, which is known as the mass-density at the space-time point (x, y, z, t).

The principle of conservation of mass says—that for an infinitely thin space-time filament, the product μdJ, where μ=mass-density at the point (x, y, z, t) of the filament (*i.e.*, the principal line of the filament), dJ=contents of the cross-section normal to the t axis, and passing through (x, y, z, t), is constant along the whole filament.

Now the contents dJ_n of the normal cross-section of the filament which is laid through (x, y, z, t) is

$$(4) \quad dJ_n = \frac{1}{\sqrt{1-u^2}} dJ = -\iota\omega_4 \; dJ = \frac{dt}{d\tau} dJ.$$

and the function $\nu = \dfrac{\mu}{-\iota\omega_4} = \mu\sqrt{1-u^2} = \mu \dfrac{\partial \tau}{\partial t}$. (5)

may be defined as the rest-mass density at the position

$(x\,y\,z\,t)$. Then the principle of conservation of mass can be formulated in this manner :—

For an infinitely thin space-time filament, the product of the rest-mass density and the contents of the normal cross-section is constant along the whole filament.

In any space-time filament, let us consider two cross-sections Q^o and Q', which have only the points on the boundary common to each other; let the space-time lines inside the filament have a larger value of t on Q' than on Q^o. The finite range enclosed between Q^o and Q' shall be called a space-time *sichel*,* Q' is the lower boundary, and Q' is the upper boundary of the *sichel.*

If we decompose a filament into elementary space-time filaments, then to an entrance-point of an elementary filament through the lower boundary of the *sichel*, there corresponds an exit point of the same by the upper boundary, whereby for both, the product νdJ_n taken in the sense of (4) and (5), has got the same value. Therefore the difference of the two integrals $\int \nu dJ_n$ (the first being extended over the upper, the second upon the lower boundary) vanishes. According to a well-known theorem of Integral Calculus the difference is equivalent to

$$\iiiint \text{lor}\ \overline{\nu\omega}\ dx\,dy\,dz\,dt,$$

the integration being extended over the whole range of the *sichel*, and (comp. (67), § 12)

$$\text{lor}\ \overline{\nu\omega} = \frac{\partial \nu\omega_1}{\partial x_1} + \frac{\partial \nu\omega_2}{\partial x_2} + \frac{\partial \nu\omega_3}{\partial x_3} + \frac{\partial \nu\omega_4}{\partial x_4}.$$

If the *sichel* reduces to a point, then the differential equation $\text{lor}\ \overline{\nu\omega} = 0$, (6)

* Sichel—a German word meaning a crescent or a scythe. The original term is retained as there is no suitable English equivalent.

which is the condition of continuity

$$\frac{\partial \mu u_x}{\partial x}+\frac{\partial \mu u_y}{\partial y}+\frac{\partial \mu u_z}{\partial z}+\frac{\partial \mu}{\partial t}=0.$$

Further let us form the integral

$$N=\iiiint \nu dx dy dz dt \tag{7}$$

extending over the whole range of the space-time *sichel*. We shall decompose the *sichel* into elementary space-time filaments, and every one of these filaments in small elements $d\tau$ of its proper-time, which are however large compared to the linear dimensions of the normal cross-section; let us assume that the mass of such a filament $\nu dJ_n=dm$ and write τ^o, τ^1 for the 'Proper-time' of the upper and lower boundary of the *sichel*.

Then the integral (7) can be denoted by

$$\iint \nu dJ_n \, d\tau=\int (\tau'-\tau^o)\, dm.$$

taken over all the elements of the sichel.

Now let us conceive of the space-time lines inside a space-time *sichel* as material curves composed of material points, and let us suppose that they are subjected to a continual change of length inside the sichel in the following manner. The entire curves are to be varied in any possible manner inside the *sichel*, while the end points on the lower and upper boundaries remain fixed, and the individual substantial points upon it are displaced in such a manner that they always move forward normal to the curves. The whole process may be analytically represented by means of a parameter λ, and to the value $\lambda=o$, shall correspond the actual curves inside the *sichel*. Such a process may be called a virtual displacement in the sichel.

Let the point (x, y, z, t) in the sichel $\lambda=o$ have the values $x+\delta x$, $y+\delta y$, $z+\delta z$, $t+\delta t$, when the parameter has the value λ; these magnitudes are then functions of (x, y, z, t, λ). Let us now conceive of an infinitely thin space-time filament at the point $(x\ y\ z\ t)$ with the normal section of contents dJ_n, and if $dJ_n + \delta dJ_n$ be the contents of the normal section at the corresponding position of the varied filament, then according to the principle of conservation of mass—($\nu + d\nu$ being the rest-mass-density at the varied position),

$$(8) \qquad (\nu+\delta\nu)\ (dJ_n + \delta dJ_n) = \nu dJ_n = dm.$$

In consequence of this condition, the integral (7) taken over the whole range of the *sichel*, varies on account of the displacement as a definite function $N+\delta N$ of λ, and we may call this function $N+\delta N$ as the *mass action* of the virtual displacement.

If we now introduce the method of writing with indices, we shall have

$$(9) \quad d(x_h + \delta x_h) = dx_h + \sum_k \frac{\partial \delta x_h}{\partial x_k} + \frac{\partial \delta x_h}{\partial \lambda}\ d\lambda$$

$k = 1, 2, 3, 4$

$h = 1, 2, 3, 4$

Now on the basis of the remarks already made, it is clear that the value of $N+\delta N$, when the value of the parameter is λ, will be :—

$$(10) \qquad N+\delta N = \iiiint \frac{\nu d(\tau+\delta\tau)}{d\tau}\ dx\ dy\ dz\ dt.$$

the integration extending over the whole sichel $d(\tau+\delta\tau)$ where $d(\tau+\delta\tau)$ denotes the magnitude, which is deduced from

$$\sqrt{-(dx_1 + d\delta x_1)^2 - (dx_2 + d\delta x_2)^2 - (dx_3 + d\delta x_3)^2 - (dx_4 + d\delta x_4)^2}$$

by means of (9) and

$$dx_1 = \omega_1\ d\tau,\ dx_2 = \omega_2\ d\tau,\ dx_3 = \omega_3\ d\tau,\ dx_4 = \omega_4\ d\tau,\ d\lambda = o$$

therefore :—

$$(11)\qquad \frac{d(\tau+\delta\tau)}{d\tau} = \sqrt{-\sum\Big(\omega_h + \sum \frac{\partial \delta x_h}{\partial x_k}\,\omega_k\Big)^2}$$

$$\left[\begin{array}{l} k=1, 2, 3, 4. \\ h=1, 2, 3, 4. \end{array}\right.$$

We shall now subject the value of the differential quotient

$$(12)\qquad \left(\frac{d(\mathrm{N}+\delta\mathrm{N})}{d\lambda}\right)_{(\lambda=0)}$$

to a transformation. Since each δx_h as a function of (x, y, z, t) vanishes for the zero-value of the paramater λ, so in general $\frac{d\delta x_k}{\partial x_h} = o$, for $\lambda = o$.

Let us now put $\left(\frac{\partial \delta x_h}{\partial \lambda}\right)_{\lambda=o} = \xi_h\ (h=1, 2, 3, 4)$ (13)

then on the basis of (10) and (11), we have the expression (12) :—

$$= -\iiiint \sum \omega_h \left(\frac{\partial \xi_h}{\partial x_1}\omega_1 + \frac{\partial \xi_h}{\partial x_2}\omega_2 + \frac{\partial \xi_h}{\partial x_3}\omega_3 + \frac{\partial \xi_h}{\partial x_4}\right)$$

$$dx\, dy\, dz\, dt$$

for the system $(x_1\ x_2\ x_3\ x_4)$ on the boundary of the *sichel*, $(\delta x_1\ \delta x_2\ \delta x_3\ \delta\ _4)$ shall vanish for every value of λ and therefore $\xi_1, \xi_2, \xi_3, \xi_4$ are nil. Then by partial integration, the integral is transformed into the form

$$\iiiint \sum \xi_h \left(\frac{\partial \nu\omega_h\omega_1}{\partial x_1} + \frac{\partial \nu\omega_h\omega_2}{\partial x_2} + \frac{\partial \nu\omega_h\omega_3}{\partial x_3} + \frac{\partial \nu\omega_h\omega_4}{\partial x_4}\right)$$

$$dx\, dy\, dz\, dt$$

the expression within the bracket may be written as

$$=\omega_h \sum \frac{\partial \nu\omega_k}{\partial x'_k} + \nu \sum \omega_k \frac{\partial \omega_h}{\partial x_k} .$$

The first sum vanishes in consequence of the continuity equation (b). The second may be written as

$$\frac{\partial \omega_h}{\partial x_1} \frac{dx_1}{d\tau} + \frac{\partial \omega_h}{\partial x_2} \frac{dx_2}{d\tau} + \frac{\partial \omega_h}{\partial x_3} \frac{dx_3}{d\tau} + \frac{\partial \omega_h}{\partial x_4} \frac{dx_4}{d\tau}$$

$$= \frac{d\omega_h}{d\tau} = \frac{d}{d\tau}\left(\frac{dx_h}{d\tau}\right)$$

whereby $\frac{d}{d\tau}$ is meant the differential quotient in the direction of the space-time line at any position. For the differential quotient (12), we obtain the final expression

$$(14) \qquad \iiiint \nu\left(\frac{\partial \omega_1}{\partial \tau}\xi_1 + \frac{\partial \omega_2}{\partial \tau}\xi_2 + \frac{\partial \omega_3}{\partial \tau}\xi_3 + \frac{\partial \omega_4}{\partial \tau}\xi_4 \right) dx\, dy\, dz\, dt.$$

For a virtual displacement in the *sichel* we have postulated the condition that the points supposed to be substantial shall advance normally to the curves giving their actual motion, which is $\lambda = o$; this condition denotes that the ξ_h is to satisfy the condition

$$w_1 \xi_1 + w_2 \xi_2 + w_3 \xi_3 + w_4 \xi_4 = o. \qquad (15)$$

Let us now turn our attention to the Maxwellian tensions in the electrodynamics of stationary bodies, and let us consider the results in § 12 and 13; then we find that Hamilton's Principle can be reconciled to the relativity postulate for continuously extended elastic media.

At every space-time point (as in § 13), let a space time matrix of the 2nd kind be known

$$(16)\ S=\begin{vmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{vmatrix} = \begin{vmatrix} X_x & Y_x & Z_x & -iT_x \\ X_y & Y_y & Z_y & -iT_y \\ X_z & Y_z & Z_z & -iT_z \\ -iX_t & -iY_t & -iZ_t & T_t \end{vmatrix}$$

where X_n Y_x .....X_x, T_t are real magnitudes.

For a virtual displacement in a space-time sichel (with the previously applied designation) the value of the integral

$$(17)\ W+\delta W=\int\int\int\int(\Sigma S_{hk}\frac{\partial(x_k+\delta x_k)}{\partial x_h}\,dx\,dy\,dz\,dt$$

extended over the whole range of the *sichel,* may be called the tensional work of the virtual displacement.

The sum which comes forth here, written in real magnitudes, is

$$X_x+Y_y+Z_z+T_t+X_x\ \frac{\partial\delta x}{\partial x}+X_y\ \frac{\partial\delta x}{\partial y}+\ldots Z_z\frac{\partial\delta z}{\partial z}$$

$$-X_t\frac{\partial\delta x}{\partial t}-\ldots+T_x\frac{\partial\delta t}{\partial x}+\ldots T_t\frac{\partial\delta t}{\partial t}$$

we can now postulate the following *minimum principle in mechanics.*

If any space-time Sichel be bounded, then for each virtual displacement in the Sichel, the sum of the mass-works, and tension works shall always be an extremum for that process of the space-time line in the Sichel which actually occurs.

The meaning is, that for each virtual displacement,

$$\left(\frac{d(\cdot\delta N+\delta W)}{d\lambda}\right)_{\lambda=0}=0 \qquad (18)$$

By applying the methods of the Calculus of Variations, the following four differential equations at once follow from this minimal principle by means of the transformation (14), and the condition (15).

$$(19) \quad \nu \frac{\partial w_h}{\partial \tau} = K_h + \chi w_h \qquad (h=1, 2, 3\ 4)$$

$$\text{whence } K_h = \frac{\partial S_{1h}}{\partial x_1} + \frac{\partial S_{2h}}{\partial x_2} + \frac{\partial S_{3h}}{\partial x_3} + \frac{\partial S_{4h}}{\partial x_4}, \qquad (20)$$

are components of the space-time vector 1st kind K = lor S, and X is a factor, which is to be determined from the relation $w\bar{w} = -1$. By multiplying (19) by w_h, and summing the four, we obtain $X = K\bar{w}$, and therefore clearly $K + (K\bar{w})w$ will be a space-time vector of the 1st kind which is normal to w. Let us write out the components of this vector as

$$X, Y, Z, iT$$

Then we arrive at the following equation for the motion of matter,

$$(21) \quad \nu \frac{d}{d\tau}\left(\frac{dx}{d\tau}\right) = X, \quad \nu \frac{d}{d\tau}\left(\frac{dy}{d\tau}\right) = Y; \quad \nu \frac{d}{d\tau}\left(\frac{dz}{d\tau}\right) = Z,$$

$\nu \frac{d}{d\tau}\left(\frac{dx}{d\tau}\right) = T$, and we have also

$$\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2 > \left(\frac{dt}{d\tau}\right)^2 = -1,$$

and $X \frac{dx}{d\tau} + Y \frac{dy}{d\tau} + Z \frac{dz}{d\tau} = T \frac{dt}{d\tau}$.

On the basis of this condition, the fourth of equations (21) is to be regarded as a direct consequence of the first three.

From (21), we can deduce the law for the motion of a material point, *i.e.*, the law for the career of an infinitely thin space-time filament.

Let $x, y, z, t,$ denote a point on a principal line chosen in any manner within the filament. We shall form the equations (21) for the points of the normal cross section of the filament through $x, y, z, t,$ and integrate them, multiplying by the elementary contents of the cross section over the whole space of the normal section. If the integrals of the right side be R_x R_y R_z R_t and if m be the constant mass of the filament, we obtain

$$(22)\quad m\frac{d}{d\tau}\frac{dx}{d\tau}=R_x,\quad m\frac{d}{d\tau}\frac{dy}{d\tau}=R_y,\quad m\frac{d}{d\tau}\frac{dz}{d\tau}=R_z,\quad m\frac{d}{d\tau}\frac{dt}{d\tau}=R_t$$

R is now a space-time vector of the 1st kind with the components (R_x R_y R_z R_t) which is normal to the space-time vector of the 1st kind w,—the velocity of the material point with the components

$$\frac{dx}{d\tau},\ \frac{dy}{d\tau},\ \frac{dz}{d\tau},\ i\ \frac{dt}{d\tau}.$$

We may call this vector R *the moving force of the material point.*

If instead of integrating over the normal section, we integrate the equations over that cross section of the filament which is normal to the t axis, and passes through (x,y,z,t), then [See (4)] the equations (22) are obtained, but are now multiplied by $\frac{d\tau}{dt}$; in particular, the last equation comes out in the form,

$$m\ \frac{d}{dt}\left(\frac{dt}{d\tau}\right)=w_x\ R_x\ \frac{d\tau}{dt}+w_y\ R_y\ \frac{d\tau}{dt}+w_z\ R_z\ \frac{d\tau}{dt}.$$

The right side is to be looked upon *as the amount of work done per unit of time* at the material point. In this equation, we obtain the energy-law for the motion of the material point and the expression

$$m\left(\frac{dt}{d\tau}-1\right)=m\left[\frac{1}{\sqrt{1-w^2}}-1\right]=m\left(\frac{1}{2}|w|^2+\frac{3}{8}|w|^4+\right)$$

may be called the kinetic energy of the material point.

Since dt is always greater than $d\tau$ we may call the quotient $\frac{dt-d\tau}{d\tau}$ as the "Gain" (vorgehen) of the time over the proper-time of the material point and the law can then be thus expressed ;—The kinetic energy of a material point is the product of its mass into the gain of the time over its proper-time.

The set of four equations (22) again shows the symmetry in (x,y,z,t), which is demanded by the relativity postulate; to the fourth equation however, a higher physical significance is to be attached, as we have already seen in the analogous case in electrodynamics. On the ground of this demand for symmetry, the triplet consisting of the first three equations are to be constructed after the model of the fourth; remembering this circumstance, we are justified in saying,—

"If the relativity-postulate be placed at the head of mechanics, then the whole set of laws of motion follows from the law of energy."

I cannot refrain from showing that no contradiction to the assumption on the relativity-postulate can be expected from the phenomena of gravitation.

If $B^*(x^*, y^*, z^*, t^*)$ be a solid (fester) space-time point, then the region of all those space-time points $B(x, y, z, t)$, for which

$$(23)\quad (x-x^*)^2+(y-y^*)^2+(z-z^*)^2=(t-t^*)^2$$

$$t-t^*\geqq 0$$

may be called a "Ray-figure" (Strahl-gebilde) of the space time point B*.

A space-time line taken in any manner can be cut by this figure only at one particular point; this easily follows from the convexity of the figure on the one hand, and on the other hand from the fact that all directions of the space-time lines are only directions from B* towards to the concave side of the figure. Then B* may be called the light-point of B.

If in (23), the point $(x\ y\ z\ t)$ be supposed to be fixed, the point $(x^*\ y^*\ z^*\ t^*)$ be supposed to be variable, then the relation (23) would represent the locus of all the space-time points B*, which are light-points of B.

Let us conceive that a material point F of mass m may, owing to the presence of another material point F*, experience a moving force according to the following law. Let us picture to ourselves the space-time filaments of F and F* along with the principal lines of the filaments. Let BC be an infinitely small element of the principal line of F; further let B* be the light point of B, C* be the light point of C on the principal line of F*; so that OA′ is the radius vector of the hyperboloidal fundamental figure (23) parallel to B*C*, finally D* is the point of intersection of line B*C* with the space normal to itself and passing through B. The moving force of the mass-point F in the space-time point B is now the space-time vector of the first kind which is normal to BC, and which is composed of the vectors

(24) $mm^*\left(\dfrac{OA'}{B^*D^*}\right)^3 BD^*$ in the direction of BD*, and

another vector of suitable value in direction of B*C*.

Now by $\left(\frac{OA'}{B^*D^*}\right)$ is to be understood the ratio of the two vectors in question. It is clear that this proposition at once shows the covariant character with respect to a Lorentz-group.

Let us now ask how the space-time filament of F behaves when the material point F* has a uniform translatory motion, *i.e.*, the principal line of the filament of F* is a line. Let us take the space time null-point in this, and by means of a Lorentz-transformation, we can take this axis as the t-axis. Let x, y, z, t, denote the point B, let τ^* denote the proper time of B*, reckoned from O. Our proposition leads to the equations

$$(25) \qquad \frac{d^2x}{d\tau^2} = -\frac{m^*x}{(t-\tau^*)^3}, \quad \frac{d^2y}{d\tau^2} = -\frac{m^*y}{(t-\tau^*)^3}.$$

$$\frac{d^2z}{d\tau^2} = \frac{-m^*z}{(t-\tau^*)^3}, \quad (26)\ \frac{d^2t}{d\tau^2} = \frac{-m^*}{(t-\tau^*)^2}\,\frac{d(t-\tau^*)}{dt}$$

where (27) $x^2+y^2+z^2=(t-\tau^*)^2$

and (28) $\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dy}{d\tau}\right)^2+\left(\frac{dz}{d\tau}\right)^2=\left(\frac{dt}{d\tau}\right)^2-1$

In consideration of (27), the three equations (25) are of the same form as the equations for the motion of a material point subjected to attraction from a fixed centre according to the Newtonian Law, only that instead of the time t, the proper time τ of the material point occurs. The fourth equation (26) gives then the connection between proper time and the time for the material point.

Now for different values of τ', the orbit of the space-point (x y z) is an ellipse with the semi-major axis a and the eccentricity e. Let E denote the excentric anomaly, T the increment of the proper time for a complete description of the orbit, finally $nT=2\pi$, so that from a properly chosen initial point τ, we have the Kepler-equation

$$(29)\qquad n\tau=E-e\sin E.$$

If we now change the unit of time, and denote the velocity of light by c, then from (28), we obtain

$$(30)\qquad \left(\frac{dt}{d\tau}\right)^2-1=\frac{m^*}{ao^2}\,\frac{1+e\cos E}{1-e\cos E}$$

Now neglecting c^{-4} with regard to 1, it follows that

$$ndt=nd\tau\left[1+\tfrac{1}{2}\,\frac{m^*}{ac^2}\,\frac{1+e\cos E}{1-e\cos E}\right]$$

from which, by applying (29),

$$(31)\qquad nt+\text{const}=\left(1+\tfrac{1}{2}\frac{m^*}{ac^2}\right)n\tau+\frac{m^*}{ac^2}\ \text{Sin}E.$$

the factor $\dfrac{m^*}{ac^2}$ is here the square of the ratio of a certain average velocity of F in its orbit to the velocity of light. If now m^* denote the mass of the sun, a the semi major axis of the earth's orbit, then this factor amounts to 10^{-8}.

The law of mass attraction which has been just described and which is formulated in accordance with the relativity postulate would signify that gravitation is propagated with the velocity of light. In view of the fact that the periodic terms in (31) are very small, it is not possible to decide out of astronomical observations between such a law (with the modified mechanics proposed above) and the Newtonian law of attraction with Newtonian mechanics.

SPACE AND TIME

A Lecture delivered before the Naturforscher Versammlung (Congress of Natural Philosophers) at Cologne—(21st September, 1908).

Gentlemen,

The conceptions about time and space, which I hope to develop before you to-day, has grown on experimental physical grounds. Herein lies its strength. The tendency is radical. Henceforth, the old conception of space for itself, and time for itself shall reduce to a mere shadow, and some sort of union of the two will be found consistent with facts.

I

Now I want to show you how we can arrive at the changed concepts about time and space from mechanics, as accepted now-a-days, from purely mathematical considerations. The equations of Newtonian mechanics show a two-fold invariance, (*i*) their form remains unaltered when we subject the fundamental space-coordinate system to any possible change of position, (*ii*) when we change the system in its nature of motion, *i. e.*, when we impress upon it any uniform motion of translation, the null-point of time plays no part. We are accustomed to look upon the axioms of geometry as settled once for all, while we seldom have the same amount of conviction regarding the axioms of mechanics, and therefore the two invariants are seldom mentioned in the same breath. Each one of these denotes a certain group of transformations for the differential equations of mechanics. We look upon the existence of the first group as a fundamental characteristics of space. We always prefer to leave off the second group to itself, and with a light heart conclude that we can never decide from physical considerations whether the space, which is supposed to be at rest, may not finally be in uniform motion. So these two groups lead quite separate existences besides each other. Their totally heterogeneous character may scare us away from the attempt to compound them. Yet it is the whole compounded group which as a whole gives us occasion for thought.

We wish to picture to ourselves the whole relation graphically. Let (x, y, z) be the rectangular coordinates of space, and t denote the time. Subjects of our perception are always connected with place and time. *No one has observed a place except at a particular time, or has observed a time except at a particular place.* Yet I respect the dogma that time and space have independent existences. I will call a space-point plus a time-point, *i.e.*, a system of values x, y, z, t, as a *world-point*. The manifoldness of all possible values of x, y, z, t, will be the *world*. I can draw four world-axes with the chalk. Now any axis drawn consists of quickly vibrating molecules, and besides, takes part in all the journeys of the earth ; and therefore gives us occasion for reflection. The greater abstraction required for the four-axes does not cause the mathematician any trouble. In order not to allow any yawning gap to exist, we shall suppose that at every place and time, something perceptible exists. In order not to specify either matter or electricity, we shall simply style these as substances. We direct our attention to the *world-point* x, y, z, t, and suppose that we are in a position to recognise this substantial point at any subsequent time. Let dt be the time element corresponding to the changes of space coordinates of this point $[dx, dy, dz]$. Then we obtain (as a picture, so to speak, of the perennial life-career of the substantial point),—a curve in the *world*—the *world-line*, the points on which unambiguously correspond to the parameter t from $+\infty$ to $-\infty$. The whole world appears to be resolved in such *world-lines*, and I may just deviate from my point if I say that according to my opinion the physical laws would find their fullest expression as mutual relations among these lines.

By this conception of time and space, the (x, y, z) manifoldness $t=o$ and its two sides $t<o$ and $t>o$ falls asunder. If for the sake of simplicity, we keep the null-point of time and space fixed, then the first named group of mechanics signifies that at $t=o$ we can give the x, y, and z-axes any possible rotation about the null-point corresponding to the homogeneous linear transformation of the expression

$$x^2+y^2+z^2.$$

The second group denotes that without changing the expression for the mechanical laws, we can substitute $(x-\alpha t, y-\beta t, z-\gamma t)$ for (x, y, z) where (α, β, γ) are any constants. According to this we can give the time-axis any possible direction in the upper half of the world $t>o$. Now what have the demands of orthogonality in space to do with this perfect freedom of the time-axis towards the upper half ?

To establish this connection, let us take a positive parameter c, and let us consider the figure

$$c^2 t^2 - x^2 - y^2 - z^2 = 1$$

According to the analogy of the hyperboloid of two sheets, this consists of two sheets separated by $t=o$. Let us consider the sheet, in the region of $t>o$, and let us now conceive the transformation of x, y, z, t in the new system of variables; (x', y', z', t') by means of which the form of the expression will remain unaltered. Clearly the rotation of space round the null-point belongs to this group of transformations. Now we can have a full idea of the transformations which we picture to ourselves from a particular transformation in which (y, z) remain unaltered. Let us draw the cross section of the upper sheets with the plane of the x- and t-axes, *i.e.*, the upper half of the hyperbola $c^2t^2 - x^2 = 1$, with its asymptotes (*vide* fig. 1).

Then let us draw the radius rector OA′, the tangent A′ B′ at A′, and let us complete the parallelogram OA′ B′ C′; also produce B′ C′ to meet the x-axis at D′. Let us now take Ox′, OA′ as new axes with the unit measuring rods $OC' = 1$, $OA' = \frac{1}{c}$; then the hyperbola is again expressed in the form $c^2t'^2 - x'^2 = 1$, $t' > o$ and the transition from (x, y, z, t) to $(x'y'z't')$ is one of the transitions in question. Let us add to this characteristic transformation any possible displacement of the space and time null-points; then we get a group of transformation depending only on c, which we may denote by G_c.

Now let us increase c to infinity. Thus $\frac{1}{c}$ becomes zero and it appears from the figure that the hyperbola is gradually shrunk into the x-axis, the asymptotic angle becomes a straight one, and every special transformation in the limit changes in such a manner that the t-axis can have any possible direction upwards, and x' more and more approximates to x. Remembering this point it is clear that the full group belonging to Newtonian Mechanics is simply the group G_c, with the value of $c = \infty$. In this state of affairs, and since G_c is mathematically more intelligible than G_∞, a mathematician may, by a free play of imagination, hit upon the thought that natural phenomena possess an invariance not only for the group G_∞, but in fact also for a group G_c, where c is finite, but yet exceedingly large compared to the usual measuring units. Such a preconception would be an extraordinary triumph for pure mathematics.

At the same time I shall remark for which value of c, this invariance can be conclusively held to be true. *For c, we shall substitute the velocity of light c in free space.* In order to avoid speaking either of space or of vacuum, we may take this quantity as the ratio between the electrostatic and electro-magnetic units of electricity.

We can form an idea of the invariant character of the expression for natural laws for the group-transformation G_c in the following manner.

Out of the totality of natural phenomena, we can, by successive higher approximations, deduce a coordinate system (x, y, z, t); by means of this coordinate system, we can represent the phenomena according to definite laws. This system of reference is by no means uniquely determined by the phenomena. *We can change the system of reference in any possible manner corresponding to the abovementioned group transformation G_c, but the expressions for natural laws will not be changed thereby.*

For example, corresponding to the above described figure, we can call t' the time, but then necessarily the space connected with it must be expressed by the manifoldness $(x' y z)$. The physical laws are now expressed by means of x', y, z, t',—and the expressions are just the same as in the case of x, y, z, t. According to this, we shall have in the world, not one space, but many spaces,—quite analogous to the case that the three-dimensional space consists of an infinite number of planes. The three-dimensional geometry will be a chapter of four-dimensional physics. Now you perceive, why I said in the beginning that time and space shall reduce to mere shadows and we shall have a world complete in itself.

II

Now the question may be asked,—what circumstances lead us to these changed views about time and space, are they not in contradiction with observed phenomena, do they finally guarantee us advantages for the description of natural phenomena?

Before we enter into the discussion, a very important point must be noticed. Suppose we have individualised time and space in any manner; then a world-line parallel to the t-axis will correspond to a stationary point; a world-line inclined to the t-axis will correspond to a point moving uniformly; and a world-curve will correspond to a point moving in any manner. Let us now picture to our mind the world-line passing through any world point x, y, z, t; now if we find the world-line parallel to the radius vector OA′ of the hyperboloidal sheet, then we can introduce OA′ as a new time-axis, and then according to the new conceptions of time and space the substance will appear to be at rest in the world point concerned. We shall now introduce this fundamental axiom:—

The substance existing at any world point can always be conceived to be at rest, if we establish our time and space suitably. The axiom denotes that in a world-point the expression

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

shall always be positive or what is equivalent to the same thing, every velocity V should be smaller than c. c shall therefore be the upper limit for all substantial velocities and herein lies a deep significance for the quantity c. At the first impression, the axiom seems to be rather unsatisfactory. It is to be remembered that only a modified mechanics will occur, in which the square root of this differential combination takes the place of time, so that cases in which the velocity is greater than c will play no part, something like imaginary coordinates in geometry.

The *impulse* and real cause of inducement *for the assumption of the group-transformation* G_c is the fact that the differential equation for the propagation of light in vacant space possesses the group-transformation G_c. On the other hand, the idea of rigid bodies has any sense only in a system mechanics with the group G_∞. Now if we have an optics with G_c, and on the other hand if there are rigid bodies, it is easy to see that a t-direction can be defined by the two hyperboloidal shells common to the groups G_∞, and G_c, which has got the further consequence, that by means of suitable rigid instruments in the laboratory, we can perceive a change in natural phenomena, in case of different orientations, with regard to the direction of progressive motion of the earth. But all efforts directed towards this object, and even the celebrated interference-experiment of Michelson have given negative results. In order to supply an explanation for this result, H. A. Lorentz formed a hypothesis which practically amounts to an invariance of optics for the group G_c. According to Lorentz every substance shall suffer a contraction $1 : \left(\sqrt{1 - \frac{v^2}{c^2}}\right)$ in length, in the direction of its motion

$$\frac{l}{l'} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \qquad l' = l\left(1 - \frac{v^2}{c^2}\right).$$

This hypothesis sounds rather phantastical. For the contraction is not to be thought of as a consequence of the resistance of ether, but purely as a gift from the skies, as a sort of condition always accompanying a state of motion.

I shall show in our figure that Lorentz's hypothesis is fully equivalent to the new conceptions about time and space. Thereby it may appear more intelligible. Let us now, for the sake of simplicity, neglect (y, z) and fix our attention on a two dimensional world, in which let upright strips parallel to the t-axis represent a state of rest and another parallel strip inclined to the t-axis represent a state of uniform motion for a body, which has a constant spatial extension (see fig. 1). If OA′ is parallel to the second strip, we can take t' as the t-axis and x' as the x-axis, then the second body will appear to be at rest, and the first body in uniform motion. We shall now assume that the first body supposed to be at rest, has the length l, *i.e.*, the cross section PP of the first strip upon the x-axis$=l\cdot$ OC, where OC is the unit measuring rod upon the x-axis—and the second body also, when supposed to be at rest, has the same length l, this means that, the cross section Q′Q′ of the second strip has a cross-section $l\cdot$OC′, when measured parallel to the x'-axis. In these two bodies, we have now images of two Lorentz-electrons, one of which is at rest and the other moves uniformly. Now if we stick to our original coordinates, then the extension of the second electron is given by the cross section QQ of the strip belonging to it measured parallel to the x-axis. Now it is clear since Q′Q′$=l\cdot$OC′, that QQ$=l\cdot$OD′.

If $\frac{dx}{dt}=v$, an easy calculation gives that

$$\text{OD}'=\text{OC}\sqrt{1-\frac{v^2}{c^2}},\quad\text{therefore}\quad\frac{\text{PP}}{\text{QQ}}=\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

This is the sense of Lorentz's hypothesis about the contraction of electrons in case of motion. On the other hand, if we conceive the second electron to be at rest, and therefore adopt the system $(x', t',)$ then the cross-section P′P′ of the strip of the electron parallel to OC′ is to be regarded as its length and we shall find the first electron shortened with reference to the second in the same proportion, for it is,

$$\frac{P'P'}{Q'Q'}=\frac{OD}{OC'}=\frac{OD'}{OC}=\frac{QQ}{PP}$$

Lorentz called the combination t' of (t and x) as the *local time* (*Ortszeit*) of the uniformly moving electron, and used a physical construction of this idea for a better comprehension of the contraction-hypothesis. But to perceive clearly that the time of an electron is as good as the time of any other electron, *i.e.* t, t' are to be regarded as equivalent, has been the service of A. Einstein [Ann. d. Phys. 891, p. 1905, Jahrb. d. Radis...4-4-11—1907] There the concept of time was shown to be completely and unambiguously established by natural phenomena. But the concept of space was not arrived at, either by Einstein or Lorentz, probably because in the case of the above-mentioned spatial transformations, where the (x', t') plane coincides with the x-t plane, the significance is possible that the x-axis of space some-how remains conserved in its position.

We can approach the idea of space in a corresponding manner, though some may regard the attempt as rather fantastical.

According to these ideas, the word "Relativity-Postulate" which has been coined for the demands of invariance in the group G, seems to be rather inexpressive for a true understanding of the group G_c, and for further progress.

Because the sense of the postulate is that the four-dimensional world is given in space and time by phenomena only, but the projection in time and space can be handled with a certain freedom, and therefore I would rather like to give to this assertion the name "*The Postulate of the Absolute world*" [World-Postulate].

III

By the world-postulate a similar treatment of the four determining quantities x, y, z, t, of a world-point is possible. Thereby the forms under which the physical laws come forth, gain in intelligibility, as I shall presently show. Above all, the idea of acceleration becomes much more striking and clear.

I shall again use the geometrical method of expression. Let us call any world-point O as a "Space-time-null-point." The cone

$$c^2t^2 - x^2 - y^2 - z^2 = O$$

consists of two parts with O as apex, one part having $t < 0'$, the other having $t > 0$. The first, which we may call *the fore-cone* consists of all those points which send light towards O, the second, which we may call *the aft-cone*. consists of all those points which receive their light from O. The region bounded by the fore-cone may be called the fore-side of O, and the region bounded by the aft-cone may be called the aft-side of O. (*Vide* fig. 2).

On the aft-side of O e have the already considered hyperboloidal shell $F = c^2t - x^2 - y^2 - z^2 = 1$, $t > O$.

The region inside the two cones will be occupied by the hyperboloid of one sheet

$$-F = x^2 + y^2 + z^2 - c^2t^2 = k^2,$$

where k^2 can have all possible positive values. The hyperbolas which lie upon this figure with O as centre, are important for us. For the sake of clearness the individual branches of this hyperbola will be called the "*Inter-hyperbola with centre O.*" Such a hyperbolic branch, when thought of as a world-line, would represent a motion which for $t = -\infty$ and $t = \infty$, asymptotically approaches the velocity of light c.

If, by way of analogy to the idea of vectors in space, we call any directed length in the manifoldness x,y,z,t a vector, then we have to distinguish between a time-vector directed from O towards the sheet $+F=1$, $t>0$ and a space-vector directed from O towards the sheet $-F=1$. The time-axis can be parallel to any vector of the first kind. Any world-point between the *fore* and *aft cones* of O, may by means of the system of reference be regarded either as synchronous with O, as well as later or earlier than O. Every world-point on the fore-side of O is necessarily always earlier, every point on the aft side of O, later than O. The limit $c = \infty$ corresponds to a complete folding up of the wedge-shaped cross-section between the fore and aft cones in the manifoldness $t = 0$. In the figure drawn, this cross-section has been intentionally drawn with a different breadth.

Let us decompose a vector drawn from O towards (x,y,z,t) into its components. If the directions of the two vectors are respectively the directions of the radius vector OR to one of the surfaces $\pm F = 1$, and of a tangent RS at the point R of the surface, then the vectors shall be called normal to each other. Accordingly

$$c^2 tt_1 - xx_1 - yy_1 - zz_1 = 0,$$

which is the condition that the vectors with the components (x, y, z, t) and $(x_1\ y_1\ z_1\ t_1)$ are normal to each other.

For the *measurement* of vectors in different directions, the unit measuring rod is to be fixed in the following manner;—a space-like vector from 0 to $-F=1$ is always to have the measure unity, and a time-like vector from O to $+F=1$, $t>0$ is always to have the measure $\frac{1}{c}$.

Let us now fix our attention upon the world-line of a substantive point running through the world-point (x, y, z, t); then as we follow the *progress* of the line, the quantity

$$d\tau = \frac{1}{c}\sqrt{c^2 dt^2 - dx^2 - dy^2 - dz^2},$$

corresponds to the time-like vector-element (dx, dy, dz, dt).

The integral $\tau = \int d\tau$, taken over the world-line from any fixed initial point P_0 to any variable final point P, may be called the " Proper-time " of the substantial point at P_0 upon the *world-line*. We may regard (x, y, z, t), *i.e.*, the components of the vector OP, as functions of the " proper-time " τ; let $(\dot{x}, \dot{y}, \dot{z}, \dot{t})$ denote the first differential-quotients, and $(\ddot{x}, \ddot{y}, \ddot{z}, \ddot{t})$ the second differential quotients of (x, y, z, t) with regard to τ, then these may respectively be called the *Velocity-vector,* and the *Acceleration-vector* of the substantial point at P. Now we have

$$\left.\begin{aligned} c^2\,\dot{t}^2 - \dot{x}^2 - \dot{y}^2 - \dot{z}^2 &= c^2 \\ c^2\,\dot{t}\,\ddot{t} - \dot{x}\,\ddot{x} - \dot{y}\,\ddot{y} - \dot{z}\,\ddot{z} &= 0 \end{aligned}\right\},$$

i.e., the '*Velocity-vector*' is the time-like vector of unit measure in the direction of the world-line at P, the '*Acceleration-vector*' at P is normal to the velocity-vector at P, and is in any case, a space-like vector.

Now there is, as can be easily seen, a certain hyperbola, which has three infinitely contiguous points in common with the world-line at P, and of which the asymptotes are the generators of a 'fore-cone' and an 'aft-cone.' This hyperbola may be called the "hyperbola of curvature" at P (*vide* fig. 3). If M be the centre of this hyperbola, then we have to deal here with an 'Inter-hyperbola' with centre M. Let P = measure of the vector MP, then we easily perceive that the acceleration-vector at P is *a vector of magnitude* $\frac{c^2}{\rho}$ *in the direction of* MP.

If $\ddot{x}$, $\ddot{y}$, $\ddot{z}$, $\ddot{t}$ are nil, then the hyperbola of curvature at P reduces to the straight line touching the world-line at P, and $\rho = \infty$.

IV

In order to demonstrate that the assumption of the group G_c for the physical laws does not possibly lead to any contradiction, it is unnecessary to undertake a revision of the whole of physics on the basis of the assumptions underlying this group. The revision has already been successfully made in the case of "Thermodynamics and

Radiation,"* for "Electromagnetic phenomena",† and finally for "Mechanics with the maintenance of the idea of mass."

For this last mentioned province of physics, the question may be asked : if there is a force with the components X, Y, Z (in the direction of the space-axes) at a world-point (x, y, z, t), where the velocity-vector is $(\dot{x}, \dot{y}, \dot{z}, \dot{t})$, then how are we to regard this force when the system of reference is changed in any possible manner? Now it is known that there are certain well-tested theorems about the ponderomotive force in electromagnetic fields, where the group G_c is undoubtedly permissible. These theorems lead us to the following simple rule; *if the system of reference be changed in any way, then the supposed force is to be put as a force in the new space-coordinates in such a manner, that the corresponding vector with the components*

$$\dot{t}X, \dot{t}Y, \dot{t}Z, \dot{t}T,$$

where $T = \frac{1}{c^2}\left(\frac{\dot{x}}{\dot{t}}X + \frac{\dot{y}}{\dot{t}}Y + \frac{\dot{z}}{\dot{t}}Z\right) = \frac{1}{c^2}$ *(the rate of which work is done at the world-point), remains unaltered.* This vector is always normal to the velocity-vector at P. Such a force-vector, representing a force at P, may be called a *moving force-vector at* P.

Now the world-line passing through P will be described by a substantial point with the constant *mechanical mass* m. Let us call *m-times* the velocity-vector at P as the

* Planck, Zur Dynamik bewegter systeme, Ann. d. physik, Bd. 26, 1908, p. 1.

† H. Minkowski; the passage refers to paper (2) of the present edition.

impulse-vector, and *m-times* the acceleration-vector at P as the *force-vector of motion*, at P. According to these definitions, the following law tells us how the motion of a point-mass takes place under any moving force-vector * :

The force-vector of motion is equal to the moving force-vector.

This enunciation comprises four equations for the components in the four directions, of which the fourth can be deduced from the first three, because both of the above-mentioned vectors are perpendicular to the velocity-vector. From the definition of T, we see that the fourth simply expresses the "Energy-law." Accordingly c^2*-times the component of the impulse-vector in the direction of the t-axis is* to be defined as *the kinetic-energy* of the point-mass. The expression for this is

$$mc^2 \frac{dt}{d\tau} = mc^2 \Big/ \sqrt{1-\frac{v^2}{c^2}}.$$

i.e., if we deduct from this the additive constant mc^2, we obtain the expression $\frac{1}{2} mv^2$ of Newtonian-mechanics upto magnitudes of *the order of* $\frac{1}{c^2}$. Hence it appears that *the energy* depends *upon the system of reference.* But since the t-axis can be laid in the direction of any time-like axis, therefore the energy-law comprises, for any possible system of reference, the whole system of equations of motion. This fact retains its significance even in the limiting case $c=\infty$, for the axiomatic construction of Newtonian mechanics, as has already been pointed out by T. R. Schütz.†

* Minkowski—Mechanics, appendix, page 65 of paper (2). Planck—Verh. d. D. P. G. Vol. 4, 1906, p. 136.

† Schütz, Gött. Nachr. 1897, p. 110.

From the very beginning, we can establish the ratio between the units of time and space in such a manner, that the velocity of light becomes unity. If we now write $\sqrt{-1}\ t = l$, in the place of l, then the differential expression

$$d\tau^2 = -(dx^2 + dy^2 + dz^2 + dl^2),$$

becomes symmetrical in (x, y, z, l); this symmetry then enters into each law, which does not contradict the *world-postulate*. We can clothe the "essential nature of this postulate in the mystical, but mathematically significant formula

$$3 \cdot 10^5 km = \sqrt{-1} \text{ Sec.}$$

V

The advantages arising from the formulation of the world-postulate are illustrated by nothing so strikingly as by the expressions which tell us about the reactions exerted by a point-charge moving in any manner according to the Maxwell-Lorentz theory.

Let us conceive of the world-line of such an electron with the charge (e), and let us introduce upon it the "Proper-time" τ reckoned from any possible initial point. In order to obtain the field caused by the electron at any world-point P_1 let us construct the fore-cone belonging to P_1 (*vide* fig. 4). Clearly this cuts the unlimited world-line of the electron at a single point P, because these directions are all time-like vectors. At P, let us draw the tangent to the world-line, and let us draw from P_1 the normal to this tangent. Let r be the measure of P_1Q. According to the definition of a fore-cone, r/c is to be reckoned as the measure of PQ. Now at the world-point P_1, the vector-potential of the field excited by e is represented by the vector in direction PQ., having the magnitude $\frac{e}{cr}$, in its three space components along the x-, y-, z-axes; the scalar-potential is represented by the component along the t-axis. This is the elementary law found out by A. Lienard, and E. Wiechert.*

If the field caused by the electron be described in the above-mentioned way, then it will appear that the division of the field into electric and magnetic forces is a relative one, and depends upon the time-axis assumed; the two forces considered together bears some analogy to the force-screw in mechanics; the analogy is, however, imperfect.

I shall now describe *the ponderomotive force which is exerted by one moving electron upon another moving electron.* Let us suppose that the world-line of a second point-electron passes through the world-point P_1. Let us determine P, Q, r as before, construct the middle-point M of the hyperbola of curvature at P, and finally the normal MN upon a line through P which is parallel to QP_1. With P as the initial point, we shall establish a system of reference in the following way: the t-axis will be laid along PQ, the x-axis in the direction of QP_1. The y-axis in the direction of MN, then the z-axis is automatically determined, as it is normal to the x-, y-, z-axes. Let $\ddot{x}, \ddot{y}, \ddot{z}, \ddot{t}$ be the acceleration-vector at P, $\dot{x}_1, \dot{y}_1, \dot{z}_1, \dot{t}_1$ be the velocity-vector at P_1. Then the force-vector exerted by the first election e, (moving in any possible manner)

* Lienard, L'Eclairage électrique T.16, 1896, p. 53.
Wiechert, Ann. d. Physik, Vol. 4.

upon the second election e, (likewise moving in any possible manner) at P_1 is represented by

$$-ee_1\left(\dot{t}_1 - \frac{\dot{x}_1}{c}\right)F,$$

For the components F_x, F_y, F_z, F_t of the vector F the following three relations hold:—

$$cF_t - F_x = \frac{1}{r^2},\ F_y = \frac{\ddot{y}}{c^2 r},\ F_z = 0,$$

and fourthly this vector F is normal to the velocity-vector P_1, and through this circumstance alone, its dependence on this last velocity-vector arises.

If we compare with this expression the previous formulæ* giving the elementary law about the ponderomotive action of moving electric charges upon each other, then we cannot but admit, that the relations which occur here reveal the inner essence of full simplicity first in four dimensions; but in three dimensions, they have very complicated projections.

In the mechanics reformed according to the world-postulate, the disharmonies which have disturbed the relations between Newtonian mechanics, and modern electrodynamics automatically disappear. I shall now consider the position of the Newtonian law of attraction to this postulate. I will assume that two point-masses m and m_1 describe their world-lines; a moving force-vector is exercised by m upon m_1, and the expression is just the same as in the case of the electron, only we have to write $+mm_1$ instead of $-ee_1$. We shall consider only the special case in which the acceleration-vector of m is always zero;

* K. Schwarzschild. Gött-Nachr. 1903.

H. A. Lorentz, Enzyklopädie der Math. Wissenschaften V. Art 14, p. 199.

then t may be introduced in such a manner that m may be regarded as fixed, the motion of m is now subjected to the moving-force vector of m alone. If we now modify this given vector by writing $\frac{\dot{1}}{\sqrt{1-\frac{v^2}{c^2}}}$ instead of $\dot{t}$ ($\dot{t}$ = 1 up to magnitudes of the order $\frac{1}{c^2}$), then it appears that Kepler's laws hold good for the position (x_1, y_1, z_1), of m_1 at any time, only in place of the time t_1, we have to write the proper time τ_1 of m_1. On the basis of this simple remark, it can be seen that the proposed law of attraction in combination with new mechanics is not less suited for the explanation of astronomical phenomena than the Newtonian law of attraction in combination with Newtonian mechanics.

Also the fundamental equations for electro-magnetic processes in moving bodies are in accordance with the world-postulate. I shall also show on a later occasion that the deduction of these equations, as taught by Lorentz, are by no means to be given up.

The fact that the world-postulate holds without exception is, I believe, the true essence of an electromagnetic picture of the world; the idea first occurred to Lorentz, its essence was first picked out by Einstein, and is now gradually fully manifest. In course of time, the mathematical consequences will be gradually deduced, and enough suggestions will be forthcoming for the experimental verification of the postulate; in this way even those, who find it uncongenial, or even painful to give up the old, time-honoured concepts, will be reconciled to the new ideas of time and space,— in the prospect that they will lead to pre-established harmony between pure mathematics and physics.

LXXIX. *The Scattering of α and β Particles by Matter and the Structure of the Atom.* *By* Professor E. RUTHERFORD, *F.R.S., University of Manchester* *.

§ 1. IT is well known that the α and β particles suffer deflexions from their rectilinear paths by encounters with atoms of matter. This scattering is far more marked for the β than for the α particle on account of the much smaller momentum and energy of the former particle. There seems to be no doubt that such swiftly moving particles pass through the atoms in their path, and that the deflexions observed are due to the strong electric field traversed within the atomic system. It has generally been supposed that the scattering of a pencil of α or β rays in passing through a thin plate of matter is the result of a multitude of small scatterings by the atoms of matter traversed. The observations, however, of Geiger and Marsden † on the scattering of α rays indicate that some of the α particles must suffer a deflexion of more than a right angle at a single encounter. They found, for example, that a small fraction of the incident α particles, about 1 in 20,000, were turned through an average angle of 90° in passing through a layer of gold-foil about ·00004 cm. thick, which was equivalent in stopping-power of the α particle to 1·6 millimetres of air. Geiger ‡ showed later that the most probable angle of deflexion for a pencil of α particles traversing a gold-foil of this thickness was about 0°·87. A simple calculation based on the theory of probability shows that the chance of an α particle being deflected through 90° is vanishingly small. In addition, it will be seen later that the distribution of the α particles for various angles of large deflexion does not follow the probability law to be expected if such large deflexions are made up of a large number of small deviations. It seems reasonable to suppose that the deflexion through a large angle is due to a single atomic encounter, for the chance of a second encounter of a kind to produce a large deflexion must in most cases be exceedingly small. A simple calculation shows that the atom must be a seat of an intense electric field in order to produce such a large deflexion at a single encounter.

Recently Sir J. J. Thomson § has put forward a theory to

* Communicated by the Author. A brief account of this paper was communicated to the Manchester Literary and Philosophical Society in February, 1911.

† Proc. Roy. Soc. lxxxii. p. 495 (1909).

‡ Proc. Roy. Soc. lxxxiii. p. 492 (1910).

§ Camb. Lit. & Phil. Soc. xv. pt. 5 (1910).

explain the scattering of electrified particles in passing through small thicknesses of matter. The atom is supposed to consist of a number N of negatively charged corpuscles, accompanied by an equal quantity of positive electricity uniformly distributed throughout a sphere. The deflexion of a negatively electrified particle in passing through the atom is ascribed to two causes—(1) the repulsion of the corpuscles distributed through the atom, and (2) the attraction of the positive electricity in the atom. The deflexion of the particle in passing through the atom is supposed to be small, while the average deflexion after a large number m of encounters was taken as $\sqrt{m}\,.\,\theta$, where θ is the average deflexion due to a single atom. It was shown that the number N of the electrons within the atom could be deduced from observations of the scattering of electrified particles. The accuracy of this theory of compound scattering was examined experimentally by Crowther * in a later paper. His results apparently confirmed the main conclusions of the theory, and he deduced, on the assumption that the positive electricity was continuous, that the number of electrons in an atom was about three times its atomic weight.

The theory of Sir J. J. Thomson is based on the assumption that the scattering due to a single atomic encounter is small, and the particular structure assumed for the atom does not admit of a very large deflexion of an α particle in traversing a single atom, unless it be supposed that the diameter of the sphere of positive electricity is minute compared with the diameter of the sphere of influence of the atom.

Since the α and β particles traverse the atom, it should be possible from a close study of the nature of the deflexion to form some idea of the constitution of the atom to produce the effects observed. In fact, the scattering of high-speed charged particles by the atoms of matter is one of the most promising methods of attack of this problem. The development of the scintillation method of counting single α particles affords unusual advantages of investigation, and the researches of H. Geiger by this method have already added much to our knowledge of the scattering of α rays by matter.

§ 2. We shall first examine theoretically the single encounters † with an atom of simple structure, which is able to

* Crowther, Proc. Roy. Soc. lxxxiv. p. 226 (1910).

† The deviation of a particle throughout a considerable angle from an encounter with a single atom will in this paper be called "single" scattering. The deviation of a particle resulting from a multitude of small deviations will be termed "compound" scattering.

produce large deflexions of an α particle, and then compare the deductions from the theory with the experimental data available.

Consider an atom which contains a charge $\pm Ne$ at its centre surrounded by a sphere of electrification containing a charge $\mp Ne$ supposed uniformly distributed throughout a sphere of radius R. e is the fundamental unit of charge, which in this paper is taken as $4{\cdot}65 \times 10^{-10}$ E.S. unit. We shall suppose that for distances less than 10^{-12} cm. the central charge and also the charge on the α particle may be supposed to be concentrated at a point. It will be shown that the main deductions from the theory are independent of whether the central charge is supposed to be positive or negative. For convenience, the sign will be assumed to be positive. The question of the stability of the atom proposed need not be considered at this stage, for this will obviously depend upon the minute structure of the atom, and on the motion of the constituent charged parts.

In order to form some idea of the forces required to deflect an α particle through a large angle, consider an atom containing a positive charge Ne at its centre, and surrounded by a distribution of negative electricity Ne uniformly distributed within a sphere of radius R. The electric force X and the potential V at a distance r from the centre of an atom for a point inside the atom, are given by

$$X = Ne\left(\frac{1}{r^2} - \frac{r}{R^3}\right)$$

$$V = Ne\left(\frac{1}{r} - \frac{3}{2R} + \frac{r^2}{2R^3}\right).$$

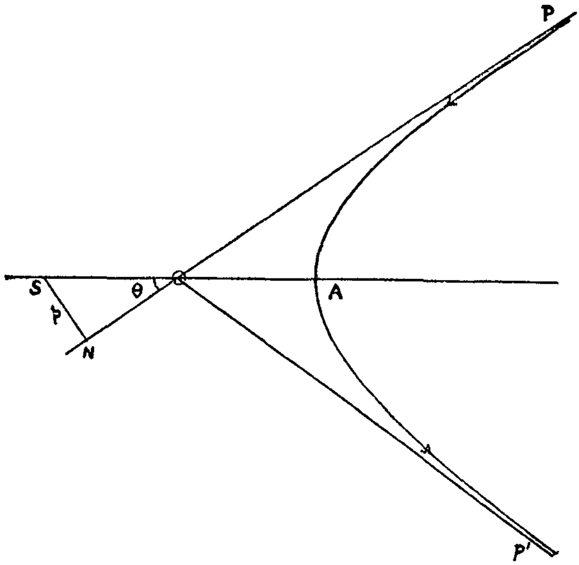
Suppose an α particle of mass m and velocity u and charge E shot directly towards the centre of the atom. It will be brought to rest at a distance b from the centre given by

$$\tfrac{1}{2}mu^2 = NeE\left(\frac{1}{b} - \frac{3}{2R} + \frac{b^2}{2R^3}\right).$$

It will be seen that b is an important quantity in later calculations. Assuming that the central charge is $100\,e$, it can be calculated that the value of b for an α particle of velocity $2{\cdot}09 \times 10^9$ cms. per second is about $3{\cdot}4 \times 10^{-12}$ cm. In this calculation b is supposed to be very small compared with R. Since R is supposed to be of the order of the radius of the atom, viz. 10^{-8} cm., it is obvious that the α particle before being turned back penetrates so close to the central charge, that the field due to the uniform distribution of negative electricity may be neglected. In general, a simple calculation shows that for all deflexions greater than a degree, we may without sensible error suppose the deflexion due to the field of the central charge alone. Possible single deviations due to the negative electricity, if distributed in the form of corpuscles, are not taken into account at this stage of the theory. It will be shown later that its effect is in general small compared with that due to the central field.

Consider the passage of a positive electrified particle close to the centre of an atom. Supposing that the velocity of the particle is not appreciably changed by its passage through the atom, the path of the particle under the influence of a repulsive force varying inversely as the square of the distance will be an hyperbola with the centre of the atom S as the external focus. Suppose the particle to enter the atom in the direction PO (fig. 1), and that the direction of motion

Fig. 1.



on escaping the atom is OP′. OP and OP′ make equal angles with the line SA, where A is the apse of the hyperbola. p = SN = perpendicular distance from centre on direction of initial motion of particle.

Let angle POA $=\theta$.

Let V = velocity of particle on entering the atom, v its velocity at A, then from consideration of angular momentum

$$pV = SA \, . \, v.$$

From conservation of energy

$$\tfrac{1}{2}mV^2 = \tfrac{1}{2}mv^2 - \frac{NeE}{SA},$$

$$v^2 = V^2\left(1 - \frac{b}{SA}\right).$$

Since the eccentricity is $\sec\theta$,

$$SA = SO + OA = p \operatorname{cosec}\theta(1+\cos\theta)$$

$$= p \cot\theta/2,$$

$$p^2 = SA(SA - b) = p\cot\theta/2(p\cot\theta/2 - b),$$

$$\therefore \quad b = 2p\cot\theta.$$

The angle of deviation ϕ of the particle is $\pi - 2\theta$ and

$$\cot\phi/2 = \frac{2p}{b}\text{*} \quad \ldots \ldots \quad (1)$$

This gives the angle of deviation of the particle in terms of b, and the perpendicular distance of the direction of projection from the centre of the atom.

For illustration, the angle of deviation ϕ for different values of p/b are shown in the following table:—

p/b	10	5	2	1	·5	·25	·125
ϕ	5°·7	11°·4	28°	53°	90°	127°	152°

§ 3. *Probability of single deflexion through any angle.*

Suppose a pencil of electrified particles to fall normally on a thin screen of matter of thickness t. With the exception of the few particles which are scattered through a large angle, the particles are supposed to pass nearly normally through the plate with only a small change of velocity. Let n = number of atoms in unit volume of material. Then the number of collisions of the particle with the atom of radius R is $\pi R^2 nt$ in the thickness t.

* A simple consideration shows that the deflexion is unaltered if the forces are attractive instead of repulsive.

The probabilty m of entering an atom within a distance p of its centre is given by

$$m=\pi p^2 nt.$$

Chance dm of striking within radii p and $p+dp$ is given by

$$dm=2\pi pnt\,.\,dp=\frac{\pi}{4}ntb^2\cot\phi/2\,\mathrm{cosec}^2\,\phi/2\,d\phi, \quad . \quad (2)$$

since $$\cot\phi/2=2p/b.$$

The value of dm gives the *fraction* of the total number of particles which are deviated between the angles ϕ and $\phi+d\phi$.

The fraction ρ of the total number of particles which are deflected through an angle greater than ϕ is given by

$$\rho=\frac{\pi}{4}ntb^2\cot^2\phi/2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The fraction ρ which is deflected between the angles ϕ_1 and ϕ_2 is given by

$$\rho=\frac{\pi}{4}ntb^2\left(\cot^2\frac{\phi_1}{2}-\cot^2\frac{\phi_2}{2}\right). \quad . \quad . \quad . \quad (4)$$

It is convenient to express the equation (2) in another form for comparison with experiment. In the case of the α rays, the number of scintillations appearing on a *constant* area of a zinc sulphide screen are counted for different angles with the direction of incidence of the particles. Let r = distance from point of incidence of α rays on scattering material, then if Q be the total number of particles falling on the scattering material, the number y of α particles falling on unit area which are deflected through an angle ϕ is given by

$$y=\frac{Qdm}{2\pi r^2\sin\phi\,.\,d\phi}=\frac{ntb^2\,.\,Q\,.\,\mathrm{cosec}^4\,\phi/2}{16r^2}. \quad . \quad . \quad (5)$$

Since $b=\dfrac{2NeE}{mu^2}$, we see from this equation that the number of α particles (scintillations) per unit area of zinc sulphide screen at a given distance r from the point of incidence of the rays is proportional to

(1) $\operatorname{cosec}^4 \phi/2$ or $1/\phi^4$ if ϕ be small;

(2) thickness of scattering material t provided this is small;

(3) magnitude of central charge Ne;

(4) and is inversely proportional to $(mu^2)^2$, or to the fourth power of the velocity if m be constant.

In these calculations, it is assumed that the α particles scattered through a large angle suffer only one large deflexion. For this to hold, it is essential that the thickness of the scattering material should be so small that the chance of a second encounter involving another large deflexion is very small. If, for example, the probability of a single deflexion ϕ in passing through a thickness t is 1/1000, the probability of two successive deflexions each of value ϕ is $1/10^6$, and is negligibly small.

The angular distribution of the α particles scattered from a thin metal sheet affords one of the simplest methods of testing the general correctness of this theory of single scattering. This has been done recently for α rays by Dr. Geiger *, who found that the distribution for particles deflected between 30° and 150° from a thin gold-foil was in substantial agreement with the theory. A more detailed account of these and other experiments to test the validity of the theory will be published later.

§ 4. *Alteration of velocity in an atomic encounter.*

It has so far been assumed that an α or β particle does not suffer an appreciable change of velocity as the result of a single atomic encounter resulting in a large deflexion of the particle. The effect of such an encounter in altering the velocity of the particle can be calculated on certain assumptions. It is supposed that only two systems are involved, viz., the swiftly moving particle and the atom which it traverses supposed initially at rest. It is supposed that the principle of conservation of momentum and of energy applies, and that there is no appreciable loss of energy or momentum by radiation.

* Manch. Lit. & Phil. Soc. 1910.

Let m be mass of the particle,

v_1 = velocity of approach,

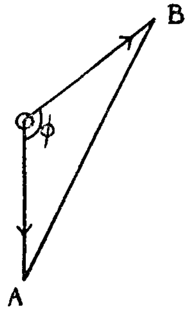
v_2 = velocity of recession,

M = mass of atom,

V = velocity communicated to atom as result of encounter.

Let OA (fig. 2) represent in magnitude and direction the momentum mv_1 of the entering particle, and OB the momentum of the receding particle which has been turned through an angle $AOB=\phi$. Then BA represents in magnitude and direction the momentum MV of the recoiling atom.

Fig. 2.



$$(MV)^2=(mv_1)^2+(mv_2)^2-2m^2v_1v_2\cos\phi. \quad (1)$$

By the conservation of energy

$$MV^2=mv_1{}^2-mv_2{}^2. \quad . \quad . \quad (2)$$

Suppose $M/m=K$ and $v_2=\rho v_1$, where ρ is <1.

From (1) and (2),

$$(K+1)\rho^2-2\rho\cos\phi=K-1,$$

or $$\rho=\frac{\cos\phi}{K+1}+\frac{1}{K+1}\sqrt{K^2-\sin^2\phi}.$$

Consider the case of an α particle of atomic weight 4, deflected through an angle of 90° by an encounter with an atom of gold of atomic weight 197.

Since $K=49$ nearly,

$$\rho=\sqrt{\frac{K-1}{K+1}}=\cdot 979,$$

or the velocity of the particle is reduced only about 2 per cent. by the encounter.

In the case of aluminium $K=27/4$ and for $\phi=90°$ $\rho=\cdot 86$.

It is seen that the reduction of velocity of the α particle becomes marked on this theory for encounters with the lighter atoms. Since the range of an α particle in air or other matter is approximately proportional to the cube of the velocity, it follows that an α particle of range 7 cms. has its range reduced to 4·5 cms. after incurring a single deviation of 90° in traversing an aluminium atom. This is of a magnitude to be easily detected experimentally. Since the value of K is very large for an encounter of a β particle with an atom, the reduction of velocity on this formula is very small.

Some very interesting cases of the theory arise in considering the changes of velocity and the distribution of scattered particles when the α particle encounters a light atom, for example a hydrogen or helium atom. A discussion of these and similar cases is reserved until the question has been examined experimentally.

§ 5. *Comparison of single and compound scattering.*

Before comparing the results of theory with experiment, it is desirable to consider the relative importance of single and compound scattering in determining the distribution of the scattered particles. Since the atom is supposed to consist of a central charge surrounded by a uniform distribution of the opposite sign through a sphere of radius R, the chance of encounters with the atom involving small deflexions is very great compared with the chance of a single large deflexion.

This question of compound scattering has been examined by Sir J. J. Thomson in the paper previously discussed (§ 1). In the notation of this paper, the average deflexion ϕ_1 due to the field of the sphere of positive electricity of radius R and quantity Ne was found by him to be

$$\phi_1 = \frac{\pi}{4} \cdot \frac{\mathrm{N}e\mathrm{E}}{mu^2} \cdot \frac{1}{\mathrm{R}}.$$

The average deflexion ϕ_2 due to the N negative corpuscles supposed distributed uniformly throughout the sphere was found to be

$$\phi_2 = \frac{16}{5} \frac{e\mathrm{E}}{mu^2} \cdot \frac{1}{\mathrm{R}} \sqrt{\frac{3\mathrm{N}}{2}}.$$

The mean deflexion due to both positive and negative electricity was taken as

$$(\phi_1^2 + \phi_2^2)^{1/2}.$$

In a similar way, it is not difficult to calculate the average deflexion due to the atom with a central charge discussed in this paper.

Since the radial electric field X at any distance r from the centre is given by

$$X=Ne\left(\frac{1}{r^2}-\frac{r}{R^3}\right),$$

it is not difficult to show that the deflexion (supposed small) of an electrified particle due to this field is given by

$$\theta=\frac{b}{p}\left(1-\frac{p^2}{R^2}\right)^{3/2},$$

where p is the perpendicular from the centre on the path of the particle and b has the same value as before. It is seen that the value of θ increases with diminution of p and becomes great for small values of ϕ.

Since we have already seen that the deflexions become very large for a particle passing near the centre of the atom, it is obviously not correct to find the average value by assuming θ is small.

Taking R of the order 10^{-8} cm., the value of p for a large deflexion is for α and β particles of the order 10^{-11} cm. Since the chance of an encounter involving a large deflexion is small compared with the chance of small deflexions, a simple consideration shows that the average small deflexion is practically unaltered if the large deflexions are omitted. This is equivalent to integrating over that part of the cross section of the atom where the deflexions are small and neglecting the small central area. It can in this way be simply shown that the average small deflexion is given by

$$\phi_1=\frac{3\pi}{8}\frac{b}{R}.$$

This value of ϕ_1 for the atom with a concentrated central charge is three times the magnitude of the average deflexion for the same value of Ne in the type of atom examined by Sir J. J. Thomson. Combining the deflexions due to the electric field and to the corpuscles, the average deflexion is

$$(\phi_1^2+\phi_2^2)^2 \quad \text{or} \quad \frac{b}{2R}\left(5\cdot54+\frac{15\cdot4}{N}\right)^{1/2}.$$

It will be seen later that the value of N is nearly proportional to the atomic weight, and is about 100 for gold. The effect due to scattering of the individual corpuscles expressed by the second term of the equation is consequently small for heavy atoms compared with that due to the distributed electric field.

Neglecting the second term, the average deflexion per atom is $\frac{3\pi b}{8R}$. We are now in a position to consider the relative effects on the distribution of particles due to single and to compound scattering. Following J. J. Thomson's argument, the average deflexion θ_t after passing through a thickness t of matter is proportional to the square root of the number of encounters and is given by

$$\theta_t = \frac{3\pi b}{8R}\sqrt{\pi R^2 . n . t} = \frac{3\pi b}{8}\sqrt{\pi n t},$$

where n as before is equal to the number of atoms per unit volume.

The probability p_1 for compound scattering that the deflexion of the particle is greater than ϕ is equal to $e^{-\phi^2/\theta_t^2}$.

Consequently $$\phi^2 = -\frac{9\pi^3}{64} b^2 nt \log p_1.$$

Next suppose that single scattering alone is operative. We have seen (§ 3) that the probability p_2 of a deflexion greater than ϕ is given by

$$p_2 = \frac{\pi}{4} b^2 . n . t \cot^2 \phi/2.$$

By comparing these two equations

$$p_2 \log p_1 = -\cdot 181 \phi^2 \cot^2 \phi/2,$$

ϕ is sufficiently small that

$$\tan \phi/2 = \phi/2,$$

$$p_2 \log p_1 = -\cdot 72.$$

If we suppose $p_2 = \cdot 5$, then $p_1 = \cdot 24$.

If $p_2 = \cdot 1$, $p_1 = \cdot 0004$.

It is evident from this comparison, that the probability for any given deflexion is always greater for single than for compound scattering. The difference is especially marked when only a small fraction of the particles are scattered through any given angle. It follows from this result that the distribution of particles due to encounters with the atoms is for small thicknesses mainly governed by single scattering. No doubt compound scattering produces some effect in equalizing the distribution of the scattered particles; but its effect becomes relatively smaller, the smaller the fraction of the particles scattered through a given angle.

§ 6. *Comparison of Theory with Experiments.*

On the present theory, the value of the central charge Ne is an important constant, and it is desirable to determine its value for different atoms. This can be most simply done by determining the small fraction of α or β particles of known velocity falling on a thin metal screen, which are scattered between ϕ and $\phi+d\phi$ where ϕ is the angle of deflexion. The influence of compound scattering should be small when this fraction is small.

Experiments in these directions are in progress, but it is desirable at this stage to discuss in the light of the present theory the data already published on scattering of α and β particles.

The following points will be discussed :—

(*a*) The "diffuse reflexion" of α particles, *i. e.* the scattering of α particles through large angles (Geiger and Marsden).

(*b*) The variation of diffuse reflexion with atomic weight of the radiator (Geiger and Marsden).

(*c*) The average scattering of a pencil of α rays transmitted through a thin metal plate (Geiger).

(*d*) The experiments of Crowther on the scattering of β rays of different velocities by various metals.

(*a*) In the paper of Geiger and Marsden (*loc. cit.*) on the diffuse reflexion of α particles falling on various substances it was shown that about 1/8000 of the α particles from radium C falling on a thick plate of platinum are scattered back in the direction of the incidence. This fraction is deduced on the assumption that the α particles are uniformly scattered in all directions, the observations being made for a deflexion of about 90°. The form of experiment is not very suited for accurate calculation, but from the data available it can be shown that the scattering observed is about that to be expected on the theory if the atom of platinum has a central charge of about 100 e.

(*b*) In their experiments on this subject, Geiger and Marsden gave the relative number of α particles diffusely reflected from thick layers of different metals, under similar conditions. The numbers obtained by them are given in the table below, where z represents the relative number of scattered particles, measured by the number of scintillations per minute on a zinc sulphide screen.

Metal.	Atomic weight.	z.	$z/A^{3/2}$.
Lead	207	62	208
Gold	197	67	242
Platinum	195	63	232
Tin	119	34	226
Silver	108	27	241
Copper	64	14·5	225
Iron	56	10·2	250
Aluminium ...	27	3·4	243
		Average	233

On the theory of single scattering, the fraction of the total number of α particles scattered through any given angle in passing through a thickness t is proportional to $n.A^2t$, assuming that the central charge is proportional to the atomic weight A. In the present case, the thickness of matter from which the scattered α particles are able to emerge and affect the zinc sulphide screen depends on the metal. Since Bragg has shown that the stopping power of an atom for an α particle is proportional to the square root of its atomic weight, the value of nt for different elements is proportional to $1/\sqrt{A}$. In this case t represents the greatest depth from which the scattered α particles emerge. The number z of α particles scattered back from a thick layer is consequently proportional to $A^{3/2}$ or $z/A^{3/2}$ should be a constant.

To compare this deduction with experiment, the relative values of the latter quotient are given in the last column. Considering the difficulty of the experiments, the agreement between theory and experiment is reasonably good *.

The single large scattering of α particles will obviously affect to some extent the shape of the Bragg ionization curve for a pencil of α rays. This effect of large scattering should be marked when the α rays have traversed screens of metals of high atomic weight, but should be small for atoms of light atomic weight.

(*c*) Geiger made a careful determination of the scattering of α particles passing through thin metal foils, by the scintillation method, and deduced the most probable angle

* The effect of change of velocity in an atomic encounter is neglected in this calculation.

through which the α particles are deflected in passing through known thicknesses of different kinds of matter.

A narrow pencil of homogeneous α rays was used as a source. After passing through the scattering foil, the total number of α particles deflected through different angles was directly measured. The angle for which the number of scattered particles was a maximum was taken as the most probable angle. The variation of the most probable angle with thickness of matter was determined, but calculation from these data is somewhat complicated by the variation of velocity of the α particles in their passage through the scattering material. A consideration of the curve of distribution of the α particles given in the paper (*loc. cit.* p. 496) shows that the angle through which half the particles are scattered is about 20 per cent greater than the most probable angle.

We have already seen that compound scattering may become important when about half the particles are scattered through a given angle, and it is difficult to disentangle in such cases the relative effects due to the two kinds of scattering. An approximate estimate can be made in the following way :—From (§ 5) the relation between the probabilities p_1 and p_2 for compound and single scattering respectively is given by

$$p_2 \log p_1 = -\cdot 721.$$

The probability q of the combined effects may as a first approximation be taken as

$$q = (p_1{}^2 + p_2{}^2)^{1/2}.$$

If $q = \cdot 5$, it follows that

$$p_1 = \cdot 2 \quad \text{and} \quad p_2 = \cdot 46.$$

We have seen that the probability p_2 of a single deflexion greater than ϕ is given by

$$p_2 = \frac{\pi}{4} n \, . \, t \, . \, b^2 \cot^2 \phi/2.$$

Since in the experiments considered ϕ is comparatively small

$$\frac{\phi \sqrt{p_2}}{\sqrt{\pi n t}} = b = \frac{2\mathrm{N}e\mathrm{E}}{mu^2}.$$

Geiger found that the most probable angle of scattering of the α rays in passing through a thickness of gold equivalent in stopping power to about ·76 cm. of air was 1° 40′. The angle ϕ through which half the α particles are turned thus corresponds to 2° nearly.

$$t = \cdot 00017 \text{ cm.} ; \; n = 6\cdot 07 \times 10^{22} ;$$

$$u \text{ (average value)} = 1\cdot 8 \times 10^9.$$

$$\mathrm{E}/m = 1\cdot 5 \times 10^{14} \text{. E.S. units} ; \; e = 4\cdot 65 \times 10^{-10}.$$

Taking the probability of single scattering $=\cdot 46$ and substituting the above values in the formula, the value of N for gold comes out to be 97.

For a thickness of gold equivalent in stopping power to 2·12 cms. of air, Geiger found the most probable angle to be 3° 40′. In this case $t=\cdot 00047$, $\phi=4^{\circ}\cdot 4$, and average $u=1\cdot 7\times 10^9$, and N comes out to be 114.

Geiger showed that the most probable angle of deflexion for an atom was nearly proportional to its atomic weight. It consequently follows that the value of N for different atoms should be nearly proportional to their atomic weights, at any rate for atomic weights between gold and aluminium.

Since the atomic weight of platinum is nearly equal to that of gold, it follows from these considerations that the magnitude of the diffuse reflexion of α particles through more than 90° from gold and the magnitude of the average small angle scattering of a pencil of rays in passing through gold-foil are both explained on the hypothesis of single scattering by supposing the atom of gold has a central charge of about $100\,e$.

(*d*) *Experiments of Crowther on scattering of β rays.*—We shall now consider how far the experimental results of Crowther on scattering of β particles of different velocities by various materials can be explained on the general theory of single scattering. On this theory, the fraction of β particles p turned through an angle greater than ϕ is given by

$$p=\frac{\pi}{4}n\,.\,t\,.\,b^2\cot^2\phi/2.$$

In most of Crowther's experiments ϕ is sufficiently small that $\tan\phi/2$ may be put equal to $\phi/2$ without much error. Consequently

$$\phi^2=2\pi\,n\,.\,t\,.\,b^2 \quad \text{if } p=1/2.$$

On the theory of compound scattering, we have already seen that the chance p_1 that the deflexion of the particles is greater than ϕ is given by

$$\phi^2/\log p_1=-\frac{9\pi^3}{64}n\,.\,t\,.\,b^2.$$

Since in the experiments of Crowther the thickness t of matter was determined for which $p_1=1/2$,

$$\phi^2=\cdot 96\pi\,n\,t\,b^2.$$

For a probability of 1/2, the theories of single and compound scattering are thus identical in general form, but differ by a numerical constant. It is thus clear that the main relations on the theory of compound scattering of Sir J. J. Thomson, which were verified experimentally by Crowther, hold equally well on the theory of single scattering.

For example, if t_m be the thickness for which half the particles are scattered through an angle ϕ, Crowther showed that $\phi/\sqrt{t_m}$ and also $\frac{mu^2}{E}\cdot\sqrt{t_m}$ were constants for a given material when ϕ was fixed. These relations hold also on the theory of single scattering. Notwithstanding this apparent similarity in form, the two theories are fundamentally different. In one case, the effects observed are due to cumulative effects of small deflexions, while in the other the large deflexions are supposed to result from a single encounter. The distribution of scattered particles is entirely different on the two theories when the probability of deflexion greater than ϕ is small.

We have already seen that the distribution of scattered α particles at various angles has been found by Geiger to be in substantial agreement with the theory of single scattering, but cannot be explained on the theory of compound scattering alone. Since there is every reason to believe that the laws of scattering of α and β particles are very similar, the law of distribution of scattered β particles should be the same as for α particles for small thicknesses of matter. Since the value of mu^2/E for the β particles is in most cases much smaller than the corresponding value for the α particles, the chance of large single deflexions for β particles in passing through a given thickness of matter is much greater than for α particles. Since on the theory of single scattering the fraction of the number of particles which are deflected through a given angle is proportional to kt, where t is the thickness supposed small and k a constant, the number of particles which are undeflected through this angle is proportional to $1-kt$. From considerations based on the theory of compound scattering, Sir J. J. Thomson deduced that the probability of deflexion less than ϕ is proportional to $1-e^{-\mu/t}$ where μ is a constant for any given value of ϕ.

The correctness of this latter formula was tested by Crowther by measuring electrically the fraction I/I_0 of the scattered β particles which passed through a circular opening subtending an angle of 36° with the scattering material. If

$$I/I_0=1-e^{-\mu/t},$$

the value of I should decrease very slowly at first with increase of t. Crowther, using aluminium as scattering material, states that the variation of I/I_0 was in good accord with this theory for small values of t. On the other hand, if single scattering be present, as it undoubtedly is for α rays, the curve showing the relation between I/I_0 and t should be nearly linear in the initial stages. The experiments of Madsen * on scattering of β rays, although not made with quite so small a thickness of aluminium as that used by Crowther, certainly support such a conclusion. Considering the importance of the point at issue, further experiments on this question are desirable.

From the table given by Crowther of the value $\phi/\sqrt{t_m}$ for different elements for β rays of velocity $2{\cdot}68 \times 10^{10}$ cms. per second, the values of the central charge Ne can be calculated on the theory of single scattering. It is supposed, as in the case of the α rays, that for the given value of $\phi/\sqrt{t_m}$ the fraction of the β particles deflected by single scattering through an angle greater than ϕ is ·46 instead of ·5.

The values of N calculated from Crowther's data are given below.

Element.	Atomic weight.	$\phi/\sqrt{t_m}$.	N.
Aluminium	27	4·25	22
Copper	63·2	10·0	42
Silver	108	15·4	78
Platinum	194	29·0	138

It will be remembered that the values of N for gold deduced from scattering of the α rays were in two calculations 97 and 114. These numbers are somewhat smaller than the values given above for platinum (viz. 138), whose atomic weight is not very different from gold. Taking into account the uncertainties involved in the calculation from the experimental data, the agreement is sufficiently close to indicate that the same general laws of scattering hold for the α and β particles, notwithstanding the wide differences in the relative velocity and mass of these particles.

As in the case of the α rays, the value of N should be most simply determined for any given element by measuring

* Phil. Mag. xviii. p. 909 (1909).

the small fraction of the incident β particles scattered through a large angle. In this way, possible errors due to small scattering will be avoided.

The scattering data for the β rays, as well as for the α rays, indicate that the central charge in an atom is approximately proportional to its atomic weight. This falls in with the experimental deductions of Schmidt *. In his theory of absorption of β rays, he supposed that in traversing a thin sheet of matter, a small fraction α of the particles are stopped, and a small fraction β are reflected or scattered back in the direction of incidence. From comparison of the absorption curves of different elements, he deduced that the value of the constant β for different elements is proportional to $n\mathrm{A}^2$ where n is the number of atoms per unit volume and A the atomic weight of the element. This is exactly the relation to be expected on the theory of single scattering if the central charge on an atom is proportional to its atomic weight.

§ 7. *General Considerations.*

In comparing the theory outlined in this paper with the experimental results, it has been supposed that the atom consists of a central charge supposed concentrated at a point, and that the large single deflexions of the α and β particles are mainly due to their passage through the strong central field. The effect of the equal and opposite compensating charge supposed distributed uniformly throughout a sphere has been neglected. Some of the evidence in support of these assumptions will now be briefly considered. For concreteness, consider the passage of a high speed α particle through an atom having a positive central charge $\mathrm{N}e$, and surrounded by a compensating charge of N electrons. Remembering that the mass, momentum, and kinetic energy of the α particle are very large compared with the corresponding values for an electron in rapid motion, it does not seem possible from dynamic considerations that an α particle can be deflected through a large angle by a close approach to an electron, even if the latter be in rapid motion and constrained by strong electrical forces. It seems reasonable to suppose that the chance of single deflexions through a large angle due to this cause, if not zero, must be exceedingly small compared with that due to the central charge.

It is of interest to examine how far the experimental evidence throws light on the question of the extent of the

* *Annal. d. Phys.* iv. 23. p. 671 (1907).

distribution of the central charge. Suppose, for example, the central charge to be composed of N unit charges distributed over such a volume that the large single deflexions are mainly due to the constituent charges and not to the external field produced by the distribution. It has been shown (§ 3) that the fraction of the α particles scattered through a large angle is proportional to $(NeE)^2$, where Ne is the central charge concentrated at a point and E the charge on the deflected particle. If, however, this charge is distributed in single units, the fraction of the α particles scattered through a given angle is proportional to Ne^2 instead of N^2e^2. In this calculation, the influence of mass of the constituent particle has been neglected, and account has only been taken of its electric field. Since it has been shown that the value of the central point charge for gold must be about 100, the value of the distributed charge required to produce the same proportion of single deflexions through a large angle should be at least 10,000. Under these conditions the mass of the constituent particle would be small compared with that of the α particle, and the difficulty arises of the production of large single deflexions at all. In addition, with such a large distributed charge, the effect of compound scattering is relatively more important than that of single scattering. For example, the probable small angle of deflexion of a pencil of α particles passing through a thin gold foil would be much greater than that experimentally observed by Geiger (§ *b–c*). The large and small angle scattering could not then be explained by the assumption of a central charge of the same value. Considering the evidence as a whole, it seems simplest to suppose that the atom contains a central charge distributed through a very small volume, and that the large single deflexions are due to the central charge as a whole, and not to its constituents. At the same time, the experimental evidence is not precise enough to negative the possibility that a small fraction of the positive charge may be carried by satellites extending some distance from the centre. Evidence on this point could be obtained by examining whether the same central charge is required to explain the large single deflexions of α and β particles; for the α particle must approach much closer to the centre of the atom than the β particle of average speed to suffer the same large deflexion.

The general data available indicate that the value of this central charge for different atoms is approximately proportional to their atomic weights, at any rate for atoms heavier than aluminium. It will be of great interest to examine experimentally whether such a simple relation holds also for the lighter atoms. In cases where the mass of the deflecting atom (for example, hydrogen, helium, lithium) is not very different from that of the α particle, the general theory of single scattering will require modification, for it is necessary to take into account the movements of the atom itself (see § 4).

It is of interest to note that Nagaoka * has mathematically considered the properties of a "Saturnian" atom which he supposed to consist of a central attracting mass surrounded by rings of rotating electrons. He showed that such a system was stable if the attractive force was large. From the point of view considered in this paper, the chance of large deflexion would practically be unaltered, whether the atom is considered to be a disk or a sphere. It may be remarked that the approximate value found for the central charge of the atom of gold ($100\,e$) is about that to be expected if the atom of gold consisted of 49 atoms of helium, each carrying a charge $2\,e$. This may be only a coincidence, but it is certainly suggestive in view of the expulsion of helium atoms carrying two unit charges from radioactive matter.

The deductions from the theory so far considered are independent of the sign of the central charge, and it has not so far been found possible to obtain definite evidence to determine whether it be positive or negative. It may be possible to settle the question of sign by consideration of the difference of the laws of absorption of the β particle to be expected on the two hypotheses, for the effect of radiation in reducing the velocity of the β particle should be far more marked with a positive than with a negative centre. If the central charge be positive, it is easily seen that a positively charged mass if released from the centre of a heavy atom, would acquire a great velocity in moving through the electric field. It may be possible in this way to account for the high velocity of expulsion of α particles without supposing that they are initially in rapid motion within the atom.

Further consideration of the application of this theory to these and other questions will be reserved for a later paper, when the main deductions of the theory have been tested experimentally. Experiments in this direction are already in progress by Geiger and Marsden.

University of Manchester,
April 1911.

* Nagaoka, Phil. Mag. vii. p. 445 (1904).

THE FOUNDATION OF THE GENERAL THEORY OF RELATIVITY

BY

A. EINSTEIN

Translated from "Die Grundlage der allgemeinen Relativitätstheorie," Annalen der Physik, 49, 1916.

THE FOUNDATION OF THE GENERAL THEORY OF RELATIVITY

By A. EINSTEIN

A. Fundamental Considerations on the Postulate of Relativity

§ 1. Observations on the Special Theory of Relativity

THE special theory of relativity is based on the following postulate, which is also satisfied by the mechanics of Galileo and Newton.

If a system of co-ordinates K is chosen so that, in relation to it, physical laws hold good in their simplest form, the *same* laws also hold good in relation to any other system of co-ordinates K′ moving in uniform translation relatively to K. This postulate we call the "special principle of relativity." The word "special" is meant to intimate that the principle is restricted to the case when K′ has a motion of uniform translation relatively to K, but that the equivalence of K′ and K does not extend to the case of non-uniform motion of K′ relatively to K.

Thus the special theory of relativity does not depart from classical mechanics through the postulate of relativity, but through the postulate of the constancy of the velocity of light *in vacuo*, from which, in combination with the special principle of relativity, there follow, in the well-known way, the relativity of simultaneity, the Lorentzian transformation, and the related laws for the behaviour of moving bodies and clocks.

The modification to which the special theory of relativity has subjected the theory of space and time is indeed far-reaching, but one important point has remained unaffected.

For the laws of geometry, even according to the special theory of relativity, are to be interpreted directly as laws relating to the possible relative positions of solid bodies at rest; and, in a more general way, the laws of kinematics are to be interpreted as laws which describe the relations of measuring bodies and clocks. To two selected material points of a stationary rigid body there always corresponds a distance of quite definite length, which is independent of the locality and orientation of the body, and is also independent of the time. To two selected positions of the hands of a clock at rest relatively to the privileged system of reference there always corresponds an interval of time of a definite length, which is independent of place and time. We shall soon see that the general theory of relativity cannot adhere to this simple physical interpretation of space and time.

§ 2. The Need for an Extension of the Postulate of Relativity

In classical mechanics, and no less in the special theory of relativity, there is an inherent epistemological defect which was, perhaps for the first time, clearly pointed out by Ernst Mach. We will elucidate it by the following example:—Two fluid bodies of the same size and nature hover freely in space at so great a distance from each other and from all other masses that only those gravitational forces need be taken into account which arise from the interaction of different parts of the same body. Let the distance between the two bodies be invariable, and in neither of the bodies let there be any relative movements of the parts with respect to one another. But let either mass, as judged by an observer at rest relatively to the other mass, rotate with constant angular velocity about the line joining the masses. This is a verifiable relative motion of the two bodies. Now let us imagine that each of the bodies has been surveyed by means of measuring instruments at rest relatively to itself, and let the surface of S_1 prove to be a sphere, and that of S_2 an ellipsoid of revolution. Thereupon we put the question—What is the reason for this difference in the two bodies? No answer can be admitted as epistemologically satisfactory,* unless the reason given is an *observable fact of experience.* The law of causality has not the significance of a statement as to the world of experience, except when *observable facts* ultimately appear as causes and effects.

Newtonian mechanics does not give a satisfactory answer to this question. It pronounces as follows :—The laws of mechanics apply to the space R_1, in respect to which the body S_1 is at rest, but not to the space R_2, in respect to which the body S_2 is at rest. But the privileged space R_1 of Galileo, thus introduced, is a merely *factitious* cause, and not a thing that can be observed. It is therefore clear that Newton's mechanics does not really satisfy the requirement of causality in the case under consideration, but only apparently does so, since it makes the factitious cause R_1 responsible for the observable difference in the bodies S_1 and S_2.

The only satisfactory answer must be that the physical system consisting of S_1 and S_2 reveals within itself no imaginable cause to which the differing behaviour of S_1 and S_2 can be referred. The cause must therefore lie *outside* this system. We have to take it that the general laws of motion, which in particular determine the shapes of S_1 and S_2, must be such that the mechanical behaviour of S_1 and S_2 is partly conditioned, in quite essential respects, by distant masses which we have not included in the system under consideration. These distant masses and their motions relative to S_1 and S_2 must then be regarded as the seat of the causes (which must be susceptible to observation) of the different behaviour of our two bodies S_1 and S_2. They take over the rôle of the factitious cause R_1. Of all imaginable spaces R_1, R_2, etc., in any kind of motion relatively to one another, there is none which we may look upon as privileged *a priori* without reviving the above-mentioned epistemological objection. *The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion.* Along this road we arrive at an extension of the postulate of relativity.

In addition to this weighty argument from the theory of

* Of course an answer may be satisfactory from the point of view of epistemology, and yet be unsound physically, if it is in conflict with other experi-

knowledge, there is a well-known physical fact which favours an extension of the theory of relativity. Let K be a Galilean system of reference, i.e. a system relatively to which (at least in the four-dimensional region under consideration) a mass, sufficiently distant from other masses, is moving with uniform motion in a straight line. Let K′ be a second system of reference which is moving relatively to K in *uniformly accelerated* translation. Then, relatively to K′, a mass sufficiently distant from other masses would have an accelerated motion such that its acceleration and direction of acceleration are independent of the material composition and physical state of the mass.

Does this permit an observer at rest relatively to K′ to infer that he is on a "really" accelerated system of reference? The answer is in the negative; for the above-mentioned relation of freely movable masses to K′ may be interpreted equally well in the following way. The system of reference K′ is unaccelerated, but the space-time territory in question is under the sway of a gravitational field, which generates the accelerated motion of the bodies relatively to K′.

This view is made possible for us by the teaching of experience as to the existence of a field of force, namely, the gravitational field, which possesses the remarkable property of imparting the same acceleration to all bodies.* The mechanical behaviour of bodies relatively to K′ is the same as presents itself to experience in the case of systems which we are wont to regard as "stationary" or as "privileged." Therefore, from the physical standpoint, the assumption readily suggests itself that the systems K and K′ may both with equal right be looked upon as "stationary," that is to say, they have an equal title as systems of reference for the physical description of phenomena.

It will be seen from these reflexions that in pursuing the general theory of relativity we shall be led to a theory of gravitation, since we are able to "produce" a gravitational field merely by changing the system of co-ordinates. It will also be obvious that the principle of the constancy of the velocity of light *in vacuo* must be modified, since we easily

* Eötvös has proved experimentally that the gravitational field has this property in great accuracy.

recognize that the path of a ray of light with respect to K′ must in general be curvilinear, if with respect to K light is propagated in a straight line with a definite constant velocity.

§ 3. The Space-Time Continuum. Requirement of General Co-Variance for the Equations Expressing General Laws of Nature

In classical mechanics, as well as in the special theory of relativity, the co-ordinates of space and time have a direct physical meaning. To say that a point-event has the X_1 co-ordinate x_1 means that the projection of the point-event on the axis of X_1, determined by rigid rods and in accordance with the rules of Euclidean geometry, is obtained by measuring off a given rod (the unit of length) x_1 times from the origin of co-ordinates along the axis of X_1. To say that a point-event has the X_4 co-ordinate $x_4 = t$, means that a standard clock, made to measure time in a definite unit period, and which is stationary relatively to the system of co-ordinates and practically coincident in space with the point-event,* will have measured off $x_4 = t$ periods at the occurrence of the event.

This view of space and time has always been in the minds of physicists, even if, as a rule, they have been unconscious of it. This is clear from the part which these concepts play in physical measurements; it must also have underlain the reader's reflexions on the preceding paragraph (§ 2) for him to connect any meaning with what he there read. But we shall now show that we must put it aside and replace it by a more general view, in order to be able to carry through the postulate of general relativity, if the special theory of relativity applies to the special case of the absence of a gravitational field.

In a space which is free of gravitational fields we introduce a Galilean system of reference K (x, y, z, t), and also a system of co-ordinates K′ (x', y', z', t') in uniform rotation relatively to K. Let the origins of both systems, as well as their axes

* We assume the possibility of verifying "simultaneity" for events immediately proximate in space, or—to speak more precisely—for immediate proximity or coincidence in space-time, without giving a definition of this fundamental concept.

of Z, permanently coincide. We shall show that for a space-time measurement in the system K′ the above definition of the physical meaning of lengths and times cannot be maintained. For reasons of symmetry it is clear that a circle around the origin in the X, Y plane of K may at the same time be regarded as a circle in the X′, Y′ plane of K′. We suppose that the circumference and diameter of this circle have been measured with a unit measure infinitely small compared with the radius, and that we have the quotient of the two results. If this experiment were performed with a measuring-rod at rest relatively to the Galilean system K, the quotient would be π. With a measuring-rod at rest relatively to K′, the quotient would be greater than π. This is readily understood if we envisage the whole process of measuring from the "stationary" system K, and take into consideration that the measuring-rod applied to the periphery undergoes a Lorentzian contraction, while the one applied along the radius does not. Hence Euclidean geometry does not apply to K′. The notion of co-ordinates defined above, which presupposes the validity of Euclidean geometry, therefore breaks down in relation to the system K′. So, too, we are unable to introduce a time corresponding to physical requirements in K′, indicated by clocks at rest relatively to K′. To convince ourselves of this impossibility, let us imagine two clocks of identical constitution placed, one at the origin of co-ordinates, and the other at the circumference of the circle, and both envisaged from the "stationary" system K. By a familiar result of the special theory of relativity, the clock at the circumference—judged from K—goes more slowly than the other, because the former is in motion and the latter at rest. An observer at the common origin of co-ordinates, capable of observing the clock at the circumference by means of light, would therefore see it lagging behind the clock beside him. As he will not make up his mind to let the velocity of light along the path in question depend explicitly on the time, he will interpret his observations as showing that the clock at the circumference "really" goes more slowly than the clock at the origin. So he will be obliged to define time in such a way that the rate of a clock depends upon where the clock may be.

We therefore reach this result:—In the general theory of relativity, space and time cannot be defined in such a way that differences of the spatial co-ordinates can be directly measured by the unit measuring-rod, or differences in the time co-ordinate by a standard clock.

The method hitherto employed for laying co-ordinates into the space-time continuum in a definite manner thus breaks down, and there seems to be no other way which would allow us to adapt systems of co-ordinates to the four-dimensional universe so that we might expect from their application a particularly simple formulation of the laws of nature. So there is nothing for it but to regard all imaginable systems of co-ordinates, on principle, as equally suitable for the description of nature. This comes to requiring that:—

The general laws of nature are to be expressed by equations which hold good for all systems of co-ordinates, that is, are co-variant with respect to any substitutions whatever (generally co-variant).

It is clear that a physical theory which satisfies this postulate will also be suitable for the general postulate of relativity. For the sum of *all* substitutions in any case includes those which correspond to all relative motions of three-dimensional systems of co-ordinates. That this requirement of general co-variance, which takes away from space and time the last remnant of physical objectivity, is a natural one, will be seen from the following reflexion. All our space-time verifications invariably amount to a determination of space-time coincidences. If, for example, events consisted merely in the motion of material points, then ultimately nothing would be observable but the meetings of two or more of these points. Moreover, the results of our measurings are nothing but verifications of such meetings of the material points of our measuring instruments with other material points, coincidences between the hands of a clock and points on the clock dial, and observed point-events happening at the same place at the same time.

The introduction of a system of reference serves no other purpose than to facilitate the description of the totality of such coincidences. We allot to the universe four space-time variables x_1, x_2, x_3, x_4 in such a way that for every point-event there is a corresponding system of values of the variables $x_1 \ldots x_4$. To two coincident point-events there corresponds one system of values of the variables $x_1 \ldots x_4$, i.e. coincidence is characterized by the identity of the co-ordinates. If, in place of the variables $x_1 \ldots x_4$, we introduce functions of them, x'_1, x'_2, x'_3, x'_4, as a new system of co-ordinates, so that the systems of values are made to correspond to one another without ambiguity, the equality of all four co-ordinates in the new system will also serve as an expression for the space-time coincidence of the two point-events. As all our physical experience can be ultimately reduced to such coincidences, there is no immediate reason for preferring certain systems of co-ordinates to others, that is to say, we arrive at the requirement of general co-variance.

§ 4. The Relation of the Four Co-ordinates to Measurement in Space and Time

It is not my purpose in this discussion to represent the general theory of relativity as a system that is as simple and logical as possible, and with the minimum number of axioms; but my main object is to develop this theory in such a way that the reader will feel that the path we have entered upon is psychologically the natural one, and that the underlying assumptions will seem to have the highest possible degree of security. With this aim in view let it now be granted that:—

For infinitely small four-dimensional regions the theory of relativity in the restricted sense is appropriate, if the co-ordinates are suitably chosen.

For this purpose we must choose the acceleration of the infinitely small ("local") system of co-ordinates so that no gravitational field occurs; this is possible for an infinitely small region. Let X_1, X_2, X_3, be the co-ordinates of space, and X_4 the appertaining co-ordinate of time measured in the appropriate unit.* If a rigid rod is imagined to be given as the unit measure, the co-ordinates, with a given orientation of the system of co-ordinates, have a direct physical meaning

* The unit of time is to be chosen so that the velocity of light *in vacuo* as measured in the "local" system of co-ordinates is to be equal to unity.

in the sense of the special theory of relativity. By the special theory of relativity the expression

$$ds^2 = -dX_1^2 - dX_2^2 - dX_3^2 + dX_4^2 \quad . \quad . \quad (1)$$

then has a value which is independent of the orientation of the local system of co-ordinates, and is ascertainable by measurements of space and time. The magnitude of the linear element pertaining to points of the four-dimensional continuum in infinite proximity, we call ds. If the ds belonging to the element $dX_1 \ldots dX_4$ is positive, we follow Minkowski in calling it time-like; if it is negative, we call it space-like.

To the "linear element" in question, or to the two infinitely proximate point-events, there will also correspond definite differentials $dx_1 \ldots dx_4$ of the four-dimensional co-ordinates of any chosen system of reference. If this system, as well as the "local" system, is given for the region under consideration, the dX_ν will allow themselves to be represented here by definite linear homogeneous expressions of the dx_σ:—

$$dX_\nu = \sum_\sigma a_{\nu\sigma} dx_\sigma \quad . \quad . \quad . \quad . \quad (2)$$

Inserting these expressions in (1), we obtain

$$ds^2 = \sum_{\tau\sigma} g_{\sigma\tau} dx_\sigma dx_\tau, \quad . \quad . \quad . \quad . \quad (3)$$

where the $g_{\sigma\tau}$ will be functions of the x_σ. These can no longer be dependent on the orientation and the state of motion of the "local" system of co-ordinates, for ds^2 is a quantity ascertainable by rod-clock measurement of point-events infinitely proximate in space-time, and defined independently of any particular choice of co-ordinates. The $g_{\sigma\tau}$ are to be chosen here so that $g_{\sigma\tau} = g_{\tau\sigma}$; the summation is to extend over all values of σ and τ, so that the sum consists of 4×4 terms, of which twelve are equal in pairs.

The case of the ordinary theory of relativity arises out of the case here considered, if it is possible, by reason of the particular relations of the $g_{\sigma\tau}$ in a finite region, to choose the system of reference in the finite region in such a way that the $g_{\sigma\tau}$ assume the constant values

$$\left.\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{matrix}\right\} \quad . \quad . \quad . \quad (4)$$

We shall find hereafter that the choice of such co-ordinates is, in general, not possible for a finite region.

From the considerations of § 2 and § 3 it follows that the quantities $g_{\tau\sigma}$ are to be regarded from the physical standpoint as the quantities which describe the gravitational field in relation to the chosen system of reference. For, if we now assume the special theory of relativity to apply to a certain four-dimensional region with the co-ordinates properly chosen, then the $g_{\sigma\tau}$ have the values given in (4). A free material point then moves, relatively to this system, with uniform motion in a straight line. Then if we introduce new space-time co-ordinates x_1, x_2, x_3, x_4, by means of any substitution we choose, the $g^{\sigma\tau}$ in this new system will no longer be constants, but functions of space and time. At the same time the motion of the free material point will present itself in the new co-ordinates as a curvilinear non-uniform motion, and the law of this motion will be independent of the nature of the moving particle. We shall therefore interpret this motion as a motion under the influence of a gravitational field. We thus find the occurrence of a gravitational field connected with a space-time variability of the g_{σ} . So, too, in the general case, when we are no longer able by a suitable choice of co-ordinates to apply the special theory of relativity to a finite region, we shall hold fast to the view that the $g_{\sigma\tau}$ describe the gravitational field.

Thus, according to the general theory of relativity, gravitation occupies an exceptional position with regard to other forces, particularly the electromagnetic forces, since the ten functions representing the gravitational field at the same time define the metrical properties of the space measured.

B. Mathematical Aids to the Formulation of Generally Covariant Equations

Having seen in the foregoing that the general postulate of relativity leads to the requirement that the equations of physics shall be covariant in the face of any substitution of the co-ordinates $x_1 \ldots x_4$, we have to consider how such generally covariant equations can be found. We now turn to this purely mathematical task, and we shall find that in its solution a fundamental rôle is played by the invariant ds given in equation (3), which, borrowing from Gauss's theory of surfaces, we have called the "linear element."

The fundamental idea of this general theory of covariants is the following:—Let certain things ("tensors") be defined with respect to any system of co-ordinates by a number of functions of the co-ordinates, called the "components" of the tensor. There are then certain rules by which these components can be calculated for a new system of co-ordinates, if they are known for the original system of co-ordinates, and if the transformation connecting the two systems is known. The things hereafter called tensors are further characterized by the fact that the equations of transformation for their components are linear and homogeneous. Accordingly, all the components in the new system vanish, if they all vanish in the original system. If, therefore, a law of nature is expressed by equating all the components of a tensor to zero, it is generally covariant. By examining the laws of the formation of tensors, we acquire the means of formulating generally covariant laws.

§ 5. Contravariant and Covariant Four-vectors

Contravariant Four-vectors.—The linear element is defined by the four "components" dx_ν, for which the law of transformation is expressed by the equation

$$dx'_\sigma = \sum_\nu \frac{\partial x'_\sigma}{\partial x_\nu} dx_\nu \quad . \quad . \quad . \quad . \quad (5)$$

The dx'_σ are expressed as linear and homogeneous functions of the dx_ν. Hence we may look upon these co-ordinate differentials as the components of a "tensor" of the particular kind which we call a contravariant four-vector. Any thing which is defined relatively to the system of co-ordinates by four quantities A^ν, and which is transformed by the same law

$$A'^\sigma = \sum_\nu \frac{\partial x'_\sigma}{\partial x_\nu} A^\nu, \quad . \quad . \quad . \quad . \quad (5a)$$

we also call a contravariant four-vector. From (5a) it follows at once that the sums $A^\sigma \pm B^\sigma$ are also components of a four-vector, if A^σ and B^σ are such. Corresponding relations hold for all "tensors" subsequently to be introduced. (Rule for the addition and subtraction of tensors.)

Covariant Four-vectors.—We call four quantities A_ν the components of a covariant four-vector, if for any arbitrary choice of the contravariant four-vector B^ν

$$\sum_\nu A_\nu B^\nu = \text{Invariant} \quad . \quad . \quad . \quad (6)$$

The law of transformation of a covariant four-vector follows from this definition. For if we replace B^ν on the right-hand side of the equation

$$\sum_\sigma A'_\sigma B'^\sigma = \sum_\nu A_\nu B^\nu$$

by the expression resulting from the inversion of (5a),

$$\sum_\sigma \frac{\partial x_\nu}{\partial x'_\sigma} B'^\sigma,$$

we obtain

$$\sum_\sigma B'^\sigma \sum_\nu \frac{\partial x_\nu}{\partial x'_\sigma} A_\nu = \sum_\sigma B'^\sigma A' \, .$$

Since this equation is true for arbitrary values of the B'^σ, it follows that the law of transformation is

$$A'_\sigma = \sum_\nu \frac{\partial x_\nu}{\partial x'_\sigma} A_\nu \quad . \quad . \quad . \quad . \quad (7)$$

Note on a Simplified Way of Writing the Expressions.—A glance at the equations of this paragraph shows that there is always a summation with respect to the indices which occur twice under a sign of summation (e.g. the index ν in (5)), and only with respect to indices which occur twice. It is therefore possible, without loss of clearness, to omit the sign of summation. In its place we introduce the convention:—If an index occurs twice in one term of an expression, it is always to be summed unless the contrary is expressly stated.

The difference between covariant and contravariant four-vectors lies in the law of transformation ((7) or (5) respectively). Both forms are tensors in the sense of the general remark above. Therein lies their importance. Following Ricci and

Levi-Civita, we denote the contravariant character by placing the index above, the covariant by placing it below.

§ 6. Tensors of the Second and Higher Ranks

Contravariant Tensors.—If we form all the sixteen products $A^{\mu\nu}$ of the components A^{μ} and B^{ν} of two contravariant four-vectors

$$A^{\mu\nu} = A^{\mu}B^{\nu} \quad . \quad . \quad . \quad . \quad (8)$$

then by (8) and (5a) $A^{\mu\nu}$ satisfies the law of transformation

$$A'^{\sigma\tau} = \frac{\partial x'_{\sigma}}{\partial x_{\mu}} \frac{\partial x'_{\tau}}{\partial x_{\nu}} A^{\mu\nu} \quad . \quad . \quad . \quad (9)$$

We call a thing which is described relatively to any system of reference by sixteen quantities, satisfying the law of transformation (9), a contravariant tensor of the second rank. Not every such tensor allows itself to be formed in accordance with (8) from two four-vectors, but it is easily shown that any given sixteen $A^{\mu\nu}$ can be represented as the sums of the $A^{\mu}B^{\nu}$ of four appropriately selected pairs of four-vectors. Hence we can prove nearly all the laws which apply to the tensor of the second rank defined by (9) in the simplest manner by demonstrating them for the special tensors of the type (8).

Contravariant Tensors of Any Rank.—It is clear that, on the lines of (8) and (9), contravariant tensors of the third and higher ranks may also be defined with 4^3 components, and so on. In the same way it follows from (8) and (9) that the contravariant four-vector may be taken in this sense as a contravariant tensor of the first rank.

Covariant Tensors.—On the other hand, if we take the sixteen products $A_{\mu\nu}$ of two covariant four-vectors A_{μ} and B_{ν},

$$A_{\mu\nu} = A_{\mu}B_{\nu}, \quad . \quad . \quad . \quad . \quad (10)$$

the law of transformation for these is

$$A'_{\sigma\tau} = \frac{\partial x_{\mu}}{\partial x'_{\sigma}} \frac{\partial x_{\nu}}{\partial x'_{\tau}} A_{\mu\nu} \quad . \quad . \quad . \quad (11)$$

This law of transformation defines the covariant tensor of the second rank. All our previous remarks on contravariant tensors apply equally to covariant tensors.

NOTE.—It is convenient to treat the scalar (or invariant) both as a contravariant and a covariant tensor of zero rank.

Mixed Tensors.—We may also define a tensor of the second rank of the type

$$A_\mu^\nu = A_\mu B^\nu \quad . \quad . \quad . \quad . \quad (12)$$

which is covariant with respect to the index μ, and contravariant with respect to the index ν. Its law of transformation is

$$A'^\tau_\sigma = \frac{\partial x'_\tau}{\partial x_\nu}\frac{\partial x_\mu}{\partial x'_\sigma}A_\mu^\nu \quad . \quad . \quad . \quad (13)$$

Naturally there are mixed tensors with any number of indices of covariant character, and any number of indices of contravariant character. Covariant and contravariant tensors may be looked upon as special cases of mixed tensors.

Symmetrical Tensors.—A contravariant, or a covariant tensor, of the second or higher rank is said to be symmetrical if two components, which are obtained the one from the other by the interchange of two indices, are equal. The tensor $A^{\mu\nu}$, or the tensor $A_{\mu\nu}$, is thus symmetrical if for any combination of the indices μ, ν,

$$A^{\mu\nu} = A^{\nu\mu}, \quad . \quad . \quad . \quad . \quad (14)$$

or respectively,

$$A_{\mu\nu} = A_{\nu\mu}. \quad . \quad . \quad . \quad . \quad (14a)$$

It has to be proved that the symmetry thus defined is a property which is independent of the system of reference. It follows in fact from (9), when (14) is taken into consideration, that

$$A'^{\sigma\tau} = \frac{\partial x'_\sigma}{\partial x_\mu}\frac{\partial x'_\tau}{\partial x_\nu}A^{\mu\nu} = \frac{\partial x'_\sigma}{\partial x_\mu}\frac{\partial x'_\tau}{\partial x_\nu}A^{\nu\mu} = \frac{\partial x'_\sigma}{\partial x_\nu}\frac{\partial x'_\tau}{\partial x_\mu}A^{\mu\nu} = A'^{\tau\sigma}.$$

The last equation but one depends upon the interchange of the summation indices μ and ν, i.e. merely on a change of notation.

Antisymmetrical Tensors.—A contravariant or a covariant tensor of the second, third, or fourth rank is said to be antisymmetrical if two components, which are obtained the one from the other by the interchange of two indices, are equal and of opposite sign. The tensor $A^{\mu\nu}$, or the tensor $A_{\mu\nu}$, is therefore antisymmetrical, if always

$$A^{\mu\nu} = -A^{\nu\mu}, \quad . \quad . \quad . \quad . \quad (15)$$

or respectively,

$$A_{\mu\nu} = -A_{\nu\mu} \quad . \quad . \quad . \quad . \quad (15a)$$

Of the sixteen components $A^{\mu\nu}$, the four components $A^{\mu\mu}$ vanish; the rest are equal and of opposite sign in pairs, so that there are only six components numerically different (a six-vector). Similarly we see that the antisymmetrical tensor of the third rank $A^{\mu\nu\sigma}$ has only four numerically different components, while the antisymmetrical tensor $A^{\mu\nu\sigma\tau}$ has only one. There are no antisymmetrical tensors of higher rank than the fourth in a continuum of four dimensions.

§ 7. Multiplication of Tensors

Outer Multiplication of Tensors.—We obtain from the components of a tensor of rank n and of a tensor of rank m the components of a tensor of rank $n + m$ by multiplying each component of the one tensor by each component of the other. Thus, for example, the tensors T arise out of the tensors A and B of different kinds,

$$T_{\mu\nu\sigma} = A_{\mu\nu}B_{\sigma},$$
$$T^{\mu\nu\sigma\tau} = A^{\mu\nu}B^{\sigma\tau},$$
$$T^{\sigma\tau}_{\mu\nu} = A_{\mu\nu}B^{\sigma\tau}.$$

The proof of the tensor character of T is given directly by the representations (8), (10), (12), or by the laws of transformation (9), (11), (13). The equations (8), (10), (12) are themselves examples of outer multiplication of tensors of the first rank.

"*Contraction*" *of a Mixed Tensor.*—From any mixed tensor we may form a tensor whose rank is less by two, by equating an index of covariant with one of contravariant character, and summing with respect to this index ("contraction"). Thus, for example, from the mixed tensor of the fourth rank $A^{\sigma\tau}_{\mu\nu}$, we obtain the mixed tensor of the second rank,

$$A^{\tau}_{\nu} = A^{\mu\tau}_{\mu\nu} \quad \left(= \sum_{\mu} A^{\mu\tau}_{\mu\nu} \right),$$

and from this, by a second contraction, the tensor of zero rank,

$$A = A^{\nu}_{\nu} = A^{\mu\nu}_{\mu\nu}.$$

The proof that the result of contraction really possesses the tensor character is given either by the representation of a tensor according to the generalization of (12) in combination with (6), or by the generalization of (13).

Inner and Mixed Multiplication of Tensors.—These consist in a combination of outer multiplication with contraction.

Examples.—From the covariant tensor of the second rank $A_{\mu\nu}$ and the contravariant tensor of the first rank B^σ we form by outer multiplication the mixed tensor

$$D_{\mu\nu}^\sigma = A_{\mu\nu}B^\sigma.$$

On contraction with respect to the indices ν and σ, we obtain the covariant four-vector

$$D_\mu = D_{\mu\nu}^\nu = A_{\mu\nu}B^\nu.$$

This we call the *inner product of the tensors* $A_{\mu\nu}$ and B^σ. Analogously we form from the tensors $A_{\mu\nu}$ and $B^{\sigma\tau}$, by outer multiplication and double contraction, the inner product $A_{\mu\nu}B^{\mu\nu}$. By outer multiplication and one contraction, we obtain from $A_{\mu\nu}$ and $B^{\sigma\tau}$ the mixed tensor of the second rank $D_\mu^\tau = A_{\mu\nu}B^{\nu\tau}$. This operation may be aptly characterized as a mixed one, being "outer" with respect to the indices μ and τ, and "inner" with respect to the indices ν and σ.

We now prove a proposition which is often useful as evidence of tensor character. From what has just been explained, $A_{\mu\nu}B^{\mu\nu}$ is a scalar if $A_{\mu\nu}$ and $B^{\sigma\tau}$ are tensors. But we may also make the following assertion: If $A_{\mu\nu}B^{\mu\nu}$ is a scalar *for any choice of the tensor* $B^{\mu\nu}$, then $A_{\mu\nu}$ has tensor character. For, by hypothesis, for any substitution,

$$A'_{\sigma\tau}B'^{\sigma\tau} = A_{\mu\nu}B^{\mu\nu}.$$

But by an inversion of (9)

$$B^{\mu\nu} = \frac{\partial x_\mu}{\partial x'_\sigma}\frac{\partial x_\nu}{\partial x'_\tau}B'^{\sigma\tau}.$$

This, inserted in the above equation, gives

$$\left(A'_{\sigma\tau} - \frac{\partial x_\mu}{\partial x'_\sigma}\frac{\partial x_\nu}{\partial x'_\tau}A_{\mu\nu}\right)B'^{\sigma\tau} = 0.$$

This can only be satisfied for arbitrary values of $B'^{\sigma\tau}$ if the bracket vanishes. The result then follows by equation (11). This rule applies correspondingly to tensors of any rank and character, and the proof is analogous in all cases.

The rule may also be demonstrated in this form: If B^μ and C^ν are any vectors, and if, for all values of these, the inner product $A_{\mu\nu}B^\mu C^\nu$ is a scalar, then $A_{\mu\nu}$ is a covariant tensor. This latter proposition also holds good even if only the more special assertion is correct, that with any choice of the four-vector B^μ the inner product $A_{\mu\nu}B^\mu B^\nu$ is a scalar, if in addition it is known that $A_{\mu\nu}$ satisfies the condition of symmetry $A_{\mu\nu} = A_{\nu\mu}$. For by the method given above we prove the tensor character of $(A_{\mu\nu} + A_{\nu\mu})$, and from this the tensor character of $A_{\mu\nu}$ follows on account of symmetry. This also can be easily generalized to the case of covariant and contravariant tensors of any rank.

Finally, there follows from what has been proved, this law, which may also be generalized for any tensors: If for any choice of the four-vector B^ν the quantities $A_{\mu\nu}B^\nu$ form a tensor of the first rank, then $A_{\mu\nu}$ is a tensor of the second rank. For, if C^μ is any four-vector, then on account of the tensor character of $A_{\mu\nu}B^\nu$, the inner product $A_{\mu\nu}B^\nu C^\mu$ is a scalar for any choice of the two four-vectors B^ν and C^μ. From which the proposition follows.

§ 8. Some Aspects of the Fundamental Tensor $g_{\mu\nu}$

The Covariant Fundamental Tensor.—In the invariant expression for the square of the linear element,

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu,$$

the part played by the dx_μ is that of a contravariant vector which may be chosen at will. Since further, $g_{\mu\nu} = g_{\nu\mu}$, it follows from the considerations of the preceding paragraph that $g_{\mu\nu}$ is a covariant tensor of the second rank. We call it the "fundamental tensor." In what follows we deduce some properties of this tensor which, it is true, apply to any tensor of the second rank. But as the fundamental tensor plays a special part in our theory, which has its physical basis in the peculiar effects of gravitation, it so happens that the relations to be developed are of importance to us only in the case of the fundamental tensor.

The Contravariant Fundamental Tensor.—If in the determinant formed by the elements $g_{\mu\nu}$, we take the co-factor of each of the $g_{\mu\nu}$ and divide it by the determinant $g = | g_{\mu\nu} |$, we obtain certain quantities $g^{\mu\nu} (= g^{\nu\mu})$ which, as we shall demonstrate, form a contravariant tensor.

By a known property of determinants

$$g_{\mu\sigma}g^{\nu\sigma} = \delta_\mu^\nu \quad . \quad . \quad . \quad . \quad (16)$$

where the symbol δ_μ^ν denotes 1 or 0, according as $\mu = \nu$ or $\mu \neq \nu$.

Instead of the above expression for ds^2 we may thus write

$$g_{\mu\sigma}\delta_\nu^\sigma dx_\mu dx_\nu$$

or, by (16)

$$g_{\mu\sigma}g_{\nu\tau}g^{\sigma\tau}dx_\mu dx_\nu.$$

But, by the multiplication rules of the preceding paragraphs, the quantities

$$d\xi_\sigma = g_{\mu\sigma}dx_\mu$$

form a covariant four-vector, and in fact an arbitrary vector, since the dx_μ are arbitrary. By introducing this into our expression we obtain

$$ds^2 = g^{\sigma\tau}d\xi_\sigma d\xi_\tau.$$

Since this, with the arbitrary choice of the vector $d\xi_\sigma$, is a scalar, and $g^{\sigma\tau}$ by its definition is symmetrical in the indices σ and τ, it follows from the results of the preceding paragraph that $g^{\sigma\tau}$ is a contravariant tensor.

It further follows from (16) that δ_μ is also a tensor, which we may call the mixed fundamental tensor.

The Determinant of the Fundamental Tensor.—By the rule for the multiplication of determinants

$$| g_{\mu\alpha}g^{\alpha\nu} | = | g_{\mu\alpha} | \times | g^{\alpha\nu} |.$$

On the other hand

$$| g_{\mu\alpha}g^{\alpha\nu} | = | \delta_\mu^\nu | = 1.$$

It therefore follows that

$$| g_{\mu\nu} | \times | g^{\mu\nu} | = 1 \quad . \quad . \quad . \quad (17)$$

The Volume Scalar.—We seek first the law of transformation of the determinant $g = | g_{\mu\nu} |$. In accordance with (11)

$$g' = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \frac{\partial x}{\partial x'_\tau} g_{\mu\nu} \right|.$$

Hence, by a double application of the rule for the multiplication of determinants, it follows that

$$g' = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \right| \cdot \left| \frac{\partial x_\nu}{\partial x'_\tau} \right| \cdot | g_{\mu\nu} | = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \right|^2 g,$$

or

$$\sqrt{g'} = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \right| \sqrt{g}.$$

On the other hand, the law of transformation of the element of volume

$$d\tau = \int dx_1 dx_2 dx_3 dx_4$$

is, in accordance with the theorem of Jacobi,

$$d\tau' = \left| \frac{\partial x'_\sigma}{\partial x_\mu} \right| d\tau.$$

By multiplication of the last two equations, we obtain

$$\sqrt{g'} d\tau' = \sqrt{g} d\tau \quad . \quad . \quad . \quad (18).$$

Instead of $\sqrt{g}$, we introduce in what follows the quantity $\sqrt{-g}$, which is always real on account of the hyperbolic character of the space-time continuum. The invariant $\sqrt{-g}d\tau$ is equal to the magnitude of the four-dimensional element of volume in the "local" system of reference, as measured with rigid rods and clocks in the sense of the special theory of relativity.

Note on the Character of the Space-time Continuum.—Our assumption that the special theory of relativity can always be applied to an infinitely small region, implies that ds^2 can always be expressed in accordance with (1) by means of real quantities $dX_1 \ . \ . \ . \ dX_4$. If we denote by $d\tau_0$ the "natural" element of volume dX_1, dX_2, dX_3, dX_4, then

$$d\tau_0 = \sqrt{-g} d\tau \quad . \quad . \quad . \quad (18a)$$

If $\sqrt{-g}$ were to vanish at a point of the four-dimensional continuum, it would mean that at this point an infinitely small "natural" volume would correspond to a finite volume in the co-ordinates. Let us assume that this is never the case. Then g cannot change sign. We will assume that, in the sense of the special theory of relativity, g always has a finite negative value. This is a hypothesis as to the physical nature of the continuum under consideration, and at the same time a convention as to the choice of co-ordinates.

But if $-g$ is always finite and positive, it is natural to settle the choice of co-ordinates *a posteriori* in such a way that this quantity is always equal to unity. We shall see later that by such a restriction of the choice of co-ordinates it is possible to achieve an important simplification of the laws of nature.

In place of (18), we then have simply $d\tau' = d\tau$, from which, in view of Jacobi's theorem, it follows that

$$\left| \frac{\partial x'_\sigma}{\partial x_\mu} \right| = 1 \quad . \quad . \quad . \quad . \quad (19)$$

Thus, with this choice of co-ordinates, only substitutions for which the determinant is unity are permissible.

But it would be erroneous to believe that this step indicates a partial abandonment of the general postulate of relativity. We do not ask "What are the laws of nature which are co-variant in face of all substitutions for which the determinant is unity?" but our question is "What are the generally co-variant laws of nature?" It is not until we have formulated these that we simplify their expression by a particular choice of the system of reference.

The Formation of New Tensors by Means of the Fundamental Tensor.—Inner, outer, and mixed multiplication of a tensor by the fundamental tensor give tensors of different character and rank. For example,

$$A^\mu = g^{\mu\sigma} A_\sigma,$$
$$A = g_{\mu\nu} A^{\mu\nu}.$$

The following forms may be specially noted:—

$$A^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} A_{\alpha\beta},$$
$$A_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} A^{\alpha\beta}$$

(the "complements" of covariant and contravariant tensors respectively), and

$$B_{\mu\nu} = g_{\mu\nu}g^{\alpha\beta}A_{\alpha\beta}.$$

We call $B_{\mu\nu}$ the reduced tensor associated with $A_{\mu\nu}$. Similarly,

$$B^{\mu\nu} = g^{\mu\nu}g_{\alpha\beta}A^{\alpha\beta}.$$

It may be noted that $g^{\mu\nu}$ is nothing more than the complement of $g_{\mu\nu}$, since

$$g^{\mu\alpha}g^{\nu\beta}g_{\alpha\beta} = g^{\mu\alpha}\delta_{\alpha}^{\nu} = g^{\mu\nu}.$$

§ 9. The Equation of the Geodetic Line. The Motion of a Particle

As the linear element ds is defined independently of the system of co-ordinates, the line drawn between two points P and P′ of the four-dimensional continuum in such a way that $\int ds$ is stationary—a geodetic line—has a meaning which also is independent of the choice of co-ordinates. Its equation is

$$\delta\int_{P}^{P'} ds = 0 \quad . \quad . \quad . \quad . \quad (20)$$

Carrying out the variation in the usual way, we obtain from this equation four differential equations which define the geodetic line; this operation will be inserted here for the sake of completeness. Let λ be a function of the co-ordinates x_ν, and let this define a family of surfaces which intersect the required geodetic line as well as all the lines in immediate proximity to it which are drawn through the points P and P′. Any such line may then be supposed to be given by expressing its co-ordinates x_ν as functions of λ. Let the symbol δ indicate the transition from a point of the required geodetic to the point corresponding to the same λ on a neighbouring line. Then for (20) we may substitute

$$\left.\begin{aligned} \int_{\lambda_1}^{\lambda_2}\delta w d\lambda &= 0 \\ w^2 &= g_{\mu\nu}\frac{dx_\mu}{d\lambda}\frac{dx_\nu}{d\lambda} \end{aligned}\right\} \quad . \quad . \quad . \quad (20a)$$

But since

$$\delta w = \frac{1}{w}\left\{\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{d\lambda}\frac{dx_\nu}{d\lambda}\delta x_\sigma + g_{\mu\nu}\frac{dx_\mu}{d\lambda}\delta\left(\frac{dx_\nu}{d\lambda}\right)\right\},$$

and

$$\delta\left(\frac{dx_\nu}{d\lambda}\right) = \frac{d}{d\lambda}(\delta x_\nu),$$

we obtain from (20a), after a partial integration,

$$\int_{\lambda_1}^{\lambda_2}\kappa_\sigma \delta x_\sigma d\lambda = 0,$$

where

$$\kappa_\sigma = \frac{d}{d\lambda}\left\{\frac{g_{\mu\nu}}{w}\frac{dx_\mu}{d\lambda}\right\} - \frac{1}{2w}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{d\lambda}\frac{dx_\nu}{d\lambda} \quad . \quad (20b)$$

Since the values of δx_σ are arbitrary, it follows from this that

$$\kappa_\sigma = 0 \quad . \quad . \quad . \quad . \quad (20c)$$

are the equations of the geodetic line.

If ds does not vanish along the geodetic line we may choose the "length of the arc" s, measured along the geodetic line, for the parameter λ. Then $w = 1$, and in place of (20c) we obtain

$$g_{\mu\nu}\frac{d^2x_\mu}{ds^2} + \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\sigma}{ds}\frac{dx_\mu}{ds} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} = 0$$

or, by a mere change of notation,

$$g_{a\sigma}\frac{d^2x_a}{ds^2} + [\mu\nu, \sigma]\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} = 0 \quad . \quad . \quad (20d)$$

where, following Christoffel, we have written

$$[\mu\nu, \sigma] = \frac{1}{2}\left(\frac{\partial g_{\mu\sigma}}{\partial x_\nu} + \frac{\partial g_{\nu\sigma}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\right) \quad . \quad . \quad (21)$$

Finally, if we multiply (20d) by $g^{\sigma\tau}$ (outer multiplication with respect to τ, inner with respect to σ), we obtain the equations of the geodetic line in the form

$$\frac{d^2x_\tau}{ds^2} + \{\mu\nu, \tau\}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} = 0 \quad . \quad . \quad . \quad (22)$$

where, following Christoffel, we have set

$$\{\mu\nu, \tau\} = g^{\tau a}[\mu\nu, a] \quad . \quad . \quad . \quad (23)$$

§ 10. The Formation of Tensors by Differentiation

With the help of the equation of the geodetic line we can now easily deduce the laws by which new tensors can be formed from old by differentiation. By this means we are able for the first time to formulate generally covariant differential equations. We reach this goal by repeated application of the following simple law :—

If in our continuum a curve is given, the points of which are specified by the arcual distance s measured from a fixed point on the curve, and if, further, ϕ is an invariant function of space, then $d\phi/ds$ is also an invariant. The proof lies in this, that ds is an invariant as well as $d\phi$.

As

$$\frac{d\phi}{ds} = \frac{\partial\phi}{\partial x_\mu}\frac{dx_\mu}{ds}$$

therefore

$$\psi = \frac{\partial\phi}{dx_\mu}\frac{dx_\mu}{ds}$$

is also an invariant, and an invariant for all curves starting from a point of the continuum, that is, for any choice of the vector dx_μ. Hence it immediately follows that

$$A_\mu = \frac{\partial\phi}{\partial x_\mu} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

is a covariant four-vector—the "gradient" of ϕ.

According to our rule, the differential quotient

$$\chi = \frac{d\psi}{ds}$$

taken on a curve, is similarly an invariant. Inserting the value of ψ, we obtain in the first place

$$\chi = \frac{\partial^2\phi}{\partial x_\mu \partial x_\nu}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} + \frac{\partial\phi}{\partial x_\mu}\frac{d^2x_\mu}{ds^2}.$$

The existence of a tensor cannot be deduced from this forthwith. But if we may take the curve along which we have differentiated to be a geodetic, we obtain on substitution for d^2x_ν/ds^2 from (22),

$$\chi = \left(\frac{\partial^2\phi}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \tau\}\frac{\partial\phi}{\partial x_\tau}\right)\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}.$$

Since we may interchange the order of the differentiations, and since by (23) and (21) $\{\mu\nu, \tau\}$ is symmetrical in μ and ν, it follows that the expression in brackets is symmetrical in μ and ν. Since a geodetic line can be drawn in any direction from a point of the continuum, and therefore dx_μ/ds is a four-vector with the ratio of its components arbitrary, it follows from the results of § 7 that

$$A_{\mu\nu} = \frac{\partial^2\phi}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \tau\}\frac{\partial\phi}{\partial x_\tau} \quad . \quad . \quad . \quad (25)$$

is a covariant tensor of the second rank. We have therefore come to this result: from the covariant tensor of the first rank

$$A_\mu = \frac{\partial\phi}{\partial x_\mu}$$

we can, by differentiation, form a covariant tensor of the second rank

$$A_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \{\mu\nu, \tau\}A_\tau \quad . \quad . \quad . \quad (26)$$

We call the tensor $A_{\mu\nu}$ the " extension " (covariant derivative) of the tensor A_μ. In the first place we can readily show that the operation leads to a tensor, even if the vector A_μ cannot be represented as a gradient. To see this, we first observe that

$$\psi\frac{\partial\phi}{\partial x_\mu}$$

is a covariant vector, if ψ and ϕ are scalars. The sum of four such terms

$$S_\mu = \psi^{(1)}\frac{\phi\partial^{(1)}}{\partial x_\mu} + . + . + \psi^{(4)}\frac{\partial\phi^{(4)}}{\partial x_\mu},$$

is also a covariant vector, if $\psi^{(1)}, \phi^{(1)} \ldots \psi^{(4)}, \phi^{(4)}$ are scalars. But it is clear that any covariant vector can be represented in the form S_μ. For, if A_μ is a vector whose components are any given functions of the x_ν, we have only to put (in terms of the selected system of co-ordinates)

$$\begin{aligned} \psi^{(1)} &= A_1, & \phi^{(1)} &= x_1, \\ \psi^{(2)} &= A_2, & \phi^{(2)} &= x_2, \\ \psi^{(3)} &= A_3, & \phi^{(3)} &= x_3, \\ \psi^{(4)} &= A_4, & \phi^{(4)} &= x_4, \end{aligned}$$

in order to ensure that S_μ shall be equal to A_μ.

Therefore, in order to demonstrate that $A_{\mu\nu}$ is a tensor if *any* covariant vector is inserted on the right-hand side for A_μ, we only need show that this is so for the vector S_μ. But for this latter purpose it is sufficient, as a glance at the right-hand side of (26) teaches us, to furnish the proof for the case

$$A_\mu = \psi \frac{\partial \phi}{\partial x_\mu}.$$

Now the right-hand side of (25) multiplied by ψ,

$$\psi \frac{\partial^2 \phi}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \tau\}\psi \frac{\partial \phi}{\partial x_\tau}$$

is a tensor. Similarly

$$\frac{\partial \psi}{\partial x_\mu} \frac{\partial \phi}{\partial x_\nu}$$

being the outer product of two vectors, is a tensor. By addition, there follows the tensor character of

$$\frac{\partial}{\partial x_\nu}\left(\psi \frac{\partial \phi}{\partial x_\mu}\right) - \{\mu\nu, \tau\}\left(\psi \frac{\partial \phi}{\partial x_\tau}\right).$$

As a glance at (26) will show, this completes the demonstration for the vector

$$\psi \frac{\partial \phi}{\partial x_\mu}$$

and consequently, from what has already been proved, for any vector A_μ.

By means of the extension of the vector, we may easily define the "extension" of a covariant tensor of any rank. This operation is a generalization of the extension of a vector. We restrict ourselves to the case of a tensor of the second rank, since this suffices to give a clear idea of the law of formation.

As has already been observed, any covariant tensor of the second rank can be represented * as the sum of tensors of the

* By outer multiplication of the vector with arbitrary components A_{11}, A_{12}, A_{13}, A_{14} by the vector with components 1, 0, 0, 0, we produce a tensor with components

$$\begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0. \end{matrix}$$

By the addition of four tensors of this type, we obtain the tensor $A_{\mu\nu}$ with any ssigned components.

type $A_\mu B_\nu$. It will therefore be sufficient to deduce the expression for the extension of a tensor of this special type. By (26) the expressions

$$\frac{\partial A_\mu}{\partial x_\sigma} - \{\sigma\mu, \tau\}A_\tau,$$

$$\frac{\partial B_\nu}{\partial x_\sigma} - \{\sigma\nu, \tau\}B_\tau,$$

are tensors. On outer multiplication of the first by B_ν, and of the second by A_μ, we obtain in each case a tensor of the third rank. By adding these, we have the tensor of the third rank

$$A_{\mu\nu\sigma} = \frac{\partial A_{\mu\nu}}{\partial x_\sigma} - \{\sigma\mu, \tau\}A_{\tau\nu} - \{\sigma\nu, \tau\}A_{\mu\tau} \quad . \quad . \quad (27)$$

where we have put $A_{\mu\nu} = A_\mu B_\nu$. As the right-hand side of (27) is linear and homogeneous in the $A_{\mu\nu}$ and their first derivatives, this law of formation leads to a tensor, not only in the case of a tensor of the type $A_\mu B_\nu$, but also in the case of a sum of such tensors, i.e. in the case of any covariant tensor of the second rank. We call $A_{\mu\nu\sigma}$ the extension of the tensor $A_{\mu\nu}$.

It is clear that (26) and (24) concern only special cases of extension (the extension of the tensors of rank one and zero respectively).

In general, all special laws of formation of tensors are included in (27) in combination with the multiplication of tensors.

§ 11. Some Cases of Special Importance

The Fundamental Tensor.—We will first prove some lemmas which will be useful hereafter. By the rule for the differentiation of determinants

$$dg = g^{\mu\nu}g\,dg_{\mu\nu} = -g_{\mu\nu}g\,dg^{\mu\nu} \quad . \quad . \quad (28)$$

The last member is obtained from the last but one, if we bear in mind that $g_{\mu\nu}g^{\mu'\nu} = \delta_\mu^{\mu'}$, so that $g_{\mu\nu}g^{\mu\nu} = 4$, and consequently

$$g_{\mu\nu}dg^{\mu\nu} + g^{\mu\nu}dg_{\mu\nu} = 0.$$

From (28), it follows that

$$\frac{1}{\sqrt{-g}}\frac{\partial\sqrt{-g}}{\partial x_\sigma} = \frac{1}{2}\frac{\partial \log(-g)}{\partial x_\sigma} = \frac{1}{2}g^{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_\sigma} = -\frac{1}{2}g_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}. \quad (29)$$

Further, from $g_{\mu\sigma}g^{\nu\sigma} = \delta^\nu_\mu$, it follows on differentiation that

$$\left.\begin{aligned} g_{\mu\sigma}dg^{\nu\sigma} &= -g^{\nu\sigma}dg_{\mu\sigma} \\ g_{\mu\sigma}\frac{\partial g^{\nu\sigma}}{\partial x_\lambda} &= -g^{\nu\sigma}\frac{\partial g_{\mu\sigma}}{\partial x_\lambda} \end{aligned}\right\} \quad . \quad . \quad . \quad (30)$$

From these, by mixed multiplication by $g^{\sigma\tau}$ and $g_{\nu\lambda}$ respectively, and a change of notation for the indices, we have

$$\left.\begin{aligned} dg^{\mu\nu} &= -g^{\mu\alpha}g^{\nu\beta}\,dg_{\alpha\beta} \\ \frac{\partial g^{\mu\nu}}{\partial x_\sigma} &= -g^{\mu\alpha}g^{\nu\beta}\frac{\partial g_{\alpha\beta}}{\partial x_\sigma} \end{aligned}\right\} \quad . \quad . \quad . \quad (31)$$

and

$$\left.\begin{aligned} dg_{\mu\nu} &= -g_{\mu\alpha}g_{\nu\beta}\,dg^{\alpha\beta} \\ \frac{\partial g_{\mu\nu}}{\partial x_\sigma} &= -g_{\mu\alpha}g_{\nu\beta}\frac{\partial g^{\alpha\beta}}{\partial x_\sigma} \end{aligned}\right\} \quad . \quad . \quad . \quad (32)$$

The relation (31) admits of a transformation, of which we also have frequently to make use. From (21)

$$\frac{\partial g_{\alpha\beta}}{\partial x_\sigma} = [\alpha\sigma, \beta] + [\beta\sigma, \alpha] \quad . \quad . \quad . \quad (33)$$

Inserting this in the second formula of (31), we obtain, in view of (23)

$$\frac{\partial g^{\mu\nu}}{\partial x_\sigma} = -g^{\mu\tau}\{\tau\sigma, \nu\} - g^{\nu\tau}\{\tau\sigma, \mu\} \quad . \quad . \quad (34)$$

Substituting the right-hand side of (34) in (29), we have

$$\frac{1}{\sqrt{-g}}\frac{\partial\sqrt{-g}}{\partial x_\sigma} = \{\mu\sigma, \mu\} \quad . \quad . \quad (29a)$$

The "Divergence" of a Contravariant Vector.—If we take the inner product of (26) by the contravariant fundamental tensor $g^{\mu\nu}$, the right-hand side, after a transformation of the first term, assumes the form

$$\frac{\partial}{\partial x_\nu}(g^{\mu\nu}A_\mu) - A_\mu\frac{\partial g^{\mu\nu}}{\partial x_\nu} - \frac{1}{2}g^{\tau\alpha}\left(\frac{\partial g_{\mu\alpha}}{\partial x_\nu} + \frac{\partial g_{\nu\alpha}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)g^{\mu\nu}A_\tau.$$

In accordance with (31) and (29), the last term of this expression may be written

$$\frac{1}{2}\frac{\partial g^{\tau\nu}}{\partial x_\nu}A_\tau + \frac{1}{2}\frac{\partial g^{\tau\mu}}{\partial x_\mu}A_\tau + \frac{1}{\sqrt{-g}}\frac{\partial\sqrt{-g}}{\partial x_\alpha}g^{\mu\nu}A_\tau.$$

As the symbols of the indices of summation are immaterial, the first two terms of this expression cancel the second of the one above. If we then write $g^{\mu\nu}A_\mu = A^\nu$, so that A^ν like A_μ is an arbitrary vector, we finally obtain

$$\Phi = \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_\nu}(\sqrt{-g}A^\nu). \quad . \quad . \quad (35)$$

This scalar is the *divergence* of the contravariant vector A^ν.

The "Curl" of a Covariant Vector.—The second term in (26) is symmetrical in the indices μ and ν. Therefore $A_{\mu\nu} - A_{\nu\mu}$ is a particularly simply constructed antisymmetrical tensor. We obtain

$$B_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A_\nu}{\partial x_\mu} \quad . \quad . \quad . \quad (36)$$

Antisymmetrical Extension of a Six-vector.—Applying (27) to an antisymmetrical tensor of the second rank $A_{\mu\nu}$, forming in addition the two equations which arise through cyclic permutations of the indices, and adding these three equations, we obtain the tensor of the third rank

$$B_{\mu\nu\sigma} = A_{\mu\nu\sigma} + A_{\nu\sigma\mu} + A_{\sigma\mu\nu} = \frac{\partial A_{\mu\nu}}{\partial x_\sigma} + \frac{\partial A_{\nu\sigma}}{\partial x_\mu} + \frac{\partial A_{\sigma\mu}}{\partial x_\nu} \quad (37)$$

which it is easy to prove is antisymmetrical.

The Divergence of a Six-vector.—Taking the mixed product of (27) by $g^{\mu\alpha}g^{\nu\beta}$, we also obtain a tensor. The first term on the right-hand side of (27) may be written in the form

$$\frac{\partial}{\partial x_\sigma}(g^{\mu\alpha}g^{\nu\beta}A_{\mu\nu}) - g^{\mu\alpha}\frac{\partial g^{\nu\beta}}{\partial x_\sigma}A_{\mu\nu} - g^{\nu\beta}\frac{\partial g^{\mu\alpha}}{\partial x_\sigma}A_{\mu\nu}.$$

If we write $A^{\alpha\beta}_\sigma$ for $g^{\mu\alpha}g^{\nu\beta}A_{\mu\nu\sigma}$ and $A^{\alpha\beta}$ for $g^{\mu\alpha}g^{\nu\beta}A_{\mu\nu}$, and in the transformed first term replace

$$\frac{\partial g^{\nu\beta}}{\partial x_\sigma} \text{ and } \frac{\partial g^{\mu\alpha}}{\partial x_\sigma}$$

by their values as given by (34), there results from the right-hand side of (27) an expression consisting of seven terms, of which four cancel, and there remains

$$A^{\alpha\beta}_{\sigma} = \frac{\partial A^{\alpha\beta}}{\partial x_{\sigma}} + \{\sigma\gamma, \alpha\}A^{\gamma\beta} + \{\sigma\gamma, \beta\}A^{\alpha\gamma}. \qquad (38)$$

This is the expression for the extension of a contravariant tensor of the second rank, and corresponding expressions for the extension of contravariant tensors of higher and lower rank may also be formed.

We note that in an analogous way we may also form the extension of a mixed tensor :—

$$A^{\alpha}_{\mu\sigma} = \frac{\partial A^{\alpha}_{\mu}}{\partial x_{\sigma}} - \{\sigma\mu, \tau\}A^{\alpha}_{\tau} + \{\sigma\tau, \alpha\}A^{\tau}_{\mu}. \qquad (39)$$

On contracting (38) with respect to the indices β and σ (inner multiplication by δ^{σ}_{β}), we obtain the vector

$$A^{\alpha} = \frac{\partial A^{\alpha\beta}}{\partial x_{\beta}} + \{\beta\gamma, \beta\}A^{\alpha\gamma} + \{\beta\gamma, \alpha\}A^{\gamma\beta}.$$

On account of the symmetry of $\{\beta\gamma, \alpha\}$ with respect to the indices β and γ, the third term on the right-hand side vanishes, if $A^{\alpha\beta}$ is, as we will assume, an antisymmetrical tensor. The second term allows itself to be transformed in accordance with (29a). Thus we obtain

$$A^{\alpha} = \frac{1}{\sqrt{-g}} \frac{\partial(\sqrt{-g}A^{\alpha\beta})}{\partial x_{\beta}}. \qquad (40)$$

This is the expression for the divergence of a contravariant six-vector.

The Divergence of a Mixed Tensor of the Second Rank.— Contracting (39) with respect to the indices α and σ, and taking (29a) into consideration, we obtain

$$\sqrt{-g}A_{\mu} = \frac{\partial(\sqrt{-g}A^{\sigma}_{\mu})}{\partial x_{\sigma}} - \{\sigma\mu, \tau\}\sqrt{-g}A^{\sigma}_{\tau}. \qquad (41)$$

If we introduce the contravariant tensor $A^{\rho\sigma} = g^{\rho\tau}A^{\sigma}_{\tau}$ in the last term, it assumes the form

$$-[\sigma\mu, \rho]\sqrt{-g}A^{\rho\sigma}.$$

If, further, the tensor $A^{\rho\sigma}$ is symmetrical, this reduces to

$$-\tfrac{1}{2}\sqrt{-g}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}A^{\rho\sigma}.$$

Had we introduced, instead of $A^{\rho\sigma}$, the covariant tensor $A_{\rho\sigma} = g_{\rho\alpha}g_{\sigma\beta}A^{\alpha\beta}$, which is also symmetrical, the last term, by virtue of (31), would assume the form

$$\tfrac{1}{2}\sqrt{-g}\frac{\partial g^{\rho\sigma}}{\partial x_\mu}A_{\rho\sigma}.$$

In the case of symmetry in question, (41) may therefore be replaced by the two forms

$$\sqrt{-g}A_\mu = \frac{\partial(\sqrt{-g}A_\mu^\sigma)}{\partial x_\sigma} - \tfrac{1}{2}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\sqrt{-g}A^{\rho\sigma} \quad . \quad (41a)$$

$$\sqrt{-g}A_\mu = \frac{\partial(\sqrt{-g}A_\mu^\sigma)}{\partial x_\sigma} + \tfrac{1}{2}\frac{\partial g^{\rho\sigma}}{\partial x_\mu}\sqrt{-g}A_{\rho\sigma} \quad . \quad (41b)$$

which we have to employ later on.

§ 12. The Riemann-Christoffel Tensor

We now seek the tensor which can be obtained from the fundamental tensor *alone*, by differentiation. At first sight the solution seems obvious. We place the fundamental tensor of the $g_{\mu\nu}$ in (27) instead of any given tensor $A_{\mu\nu}$, and thus have a new tensor, namely, the extension of the fundamental tensor. But we easily convince ourselves that this extension vanishes identically. We reach our goal, however, in the following way. In (27) place

$$A_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \{\mu\nu, \rho\}A_\rho,$$

i.e. the extension of the four-vector A_μ. Then (with a somewhat different naming of the indices) we get the tensor of the third rank

$$A_{\mu\sigma\tau} = \frac{\partial^2 A_\mu}{\partial x_\sigma \partial x_\tau} - \{\mu\sigma, \rho\}\frac{\partial A_\rho}{\partial x_\tau} - \{\mu\tau, \rho\}\frac{\partial A_\rho}{\partial x_\sigma} - \{\sigma\tau, \rho\}\frac{\partial A_\mu}{\partial x_\rho}$$
$$+ \left[-\frac{\partial}{\partial x_\tau}\{\mu\sigma, \rho\} + \{\mu\tau, \alpha\}\{\alpha\sigma, \rho\} + \{\sigma\tau, \alpha\}\{\alpha\mu, \rho\}\right]A_\rho.$$

This expression suggests forming the tensor $A_{\mu\sigma\tau} - A_{\mu\tau\sigma}$. For, if we do so, the following terms of the expression for $A_{\mu\sigma\tau}$ cancel those of $A_{\mu\tau\sigma}$, the first, the fourth, and the member corresponding to the last term in square brackets; because all these are symmetrical in σ and τ. The same holds good for the sum of the second and third terms. Thus we obtain

$$A_{\mu\sigma\tau} - A_{\mu\tau\sigma} = B^{\rho}_{\mu\sigma\tau}A_{\rho} \quad . \quad . \quad . \quad (42)$$

where

$$B^{\rho}_{\mu\sigma\tau} = -\frac{\partial}{\partial x_{\tau}}\{\mu\sigma, \rho\} + \frac{\partial}{\partial x_{\sigma}}\{\mu\tau, \rho\} - \{\mu\sigma, \alpha\}\{\alpha\tau, \rho\}$$

$$+ \{\mu\tau, \alpha\}\{\alpha\sigma, \rho\} \quad (43)$$

The essential feature of the result is that on the right side of (42) the A_{ρ} occur alone, without their derivatives. From the tensor character of $A_{\mu\sigma\tau} - A_{\mu\tau\sigma}$ in conjunction with the fact that A_{ρ} is an arbitrary vector, it follows, by reason of § 7, that $B^{\rho}_{\mu\sigma\tau}$ is a tensor (the Riemann-Christoffel tensor).

The mathematical importance of this tensor is as follows: If the continuum is of such a nature that there is a co-ordinate system with reference to which the $g_{\mu\nu}$ are constants, then all the $B^{\rho}_{\mu\sigma\tau}$ vanish. If we choose any new system of co-ordinates in place of the original ones, the $g_{\mu\nu}$ referred thereto will not be constants, but in consequence of its tensor nature, the transformed components of $B^{\rho}_{\mu\sigma\tau}$ will still vanish in the new system. Thus the vanishing of the Riemann tensor is a necessary condition that, by an appropriate choice of the system of reference, the $g_{\mu\nu}$ may be constants. In our problem this corresponds to the case in which,* with a suitable choice of the system of reference, the special theory of relativity holds good for a *finite* region of the continuum.

Contracting (43) with respect to the indices τ and ρ we obtain the covariant tensor of second rank

* The mathematicians have proved that this is also a *sufficient* condition.

$$\left.\begin{aligned} G_{\mu\nu} &= B^{\rho}_{\mu\nu\rho} = R_{\mu\nu} + S_{\mu\nu} \\ R_{\mu\nu} &= -\frac{\partial}{\partial x_{\alpha}}\{\mu\nu, \alpha\} + \{\mu\alpha, \beta\}\{\nu\beta, \alpha\} \\ S_{\mu\nu} &= \frac{\partial^2 \log\sqrt{-g}}{\partial x_{\mu}\partial x_{\nu}} - \{\mu\nu, \alpha\}\frac{\partial \log\sqrt{-g}}{\partial x_{\alpha}} \end{aligned}\right\} \quad (44)$$

where

Note on the Choice of Co-ordinates.—It has already been observed in § 8, in connexion with equation (18a), that the choice of co-ordinates may with advantage be made so that $\sqrt{-g} = 1$. A glance at the equations obtained in the last two sections shows that by such a choice the laws of formation of tensors undergo an important simplification. This applies particularly to $G_{\mu\nu}$, the tensor just developed, which plays a fundamental part in the theory to be set forth. For this specialization of the choice of co-ordinates brings about the vanishing of $S_{\mu\nu}$, so that the tensor $G_{\mu\nu}$ reduces to $R_{\mu\nu}$.

On this account I shall hereafter give all relations in the simplified form which this specialization of the choice of co-ordinates brings with it. It will then be an easy matter to revert to the *generally* covariant equations, if this seems desirable in a special case.

C. Theory of the Gravitational Field

§ 13. Equations of Motion of a Material Point in the Gravitational Field. Expression for the Field-components of Gravitation

A freely movable body not subjected to external forces moves, according to the special theory of relativity, in a straight line and uniformly. This is also the case, according to the general theory of relativity, for a part of four-dimensional space in which the system of co-ordinates K_0, may be, and is, so chosen that they have the special constant values given in (4).

If we consider precisely this movement from any chosen system of co-ordinates K_1, the body, observed from K_1, moves, according to the considerations in § 2, in a gravitational field. The law of motion with respect to K_1 results without difficulty from the following consideration. With respect to K_0 the law of motion corresponds to a four-dimensional straight line, i.e. to a geodetic line. Now since the geodetic line is defined independently of the system of reference, its equations will also be the equation of motion of the material point with respect to K_1. If we set

$$\Gamma^{\tau}_{\mu\nu} = -\{\mu\nu, \tau\} \quad . \quad . \quad . \quad (45)$$

the equation of the motion of the point with respect to K_1, becomes

$$\frac{d^2 x_\tau}{ds^2} = \Gamma^{\tau}_{\mu\nu} \frac{dx_\mu}{ds} \frac{dx_\nu}{ds} \quad . \quad . \quad . \quad (46)$$

We now make the assumption, which readily suggests itself, that this covariant system of equations also defines the motion of the point in the gravitational field in the case when there is no system of reference K_0, with respect to which the special theory of relativity holds good in a finite region. We have all the more justification for this assumption as (46) contains only *first* derivatives of the $g_{\mu\nu}$, between which even in the special case of the existence of K_0, no relations subsist.*

If the $\Gamma^{\tau}_{\mu\nu}$ vanish, then the point moves uniformly in a straight line. These quantities therefore condition the deviation of the motion from uniformity. They are the components of the gravitational field.

§ 14. The Field Equations of Gravitation in the Absence of Matter

We make a distinction hereafter between "gravitational field" and "matter" in this way, that we denote everything but the gravitational field as "matter." Our use of the word therefore includes not only matter in the ordinary sense, but the electromagnetic field as well.

Our next task is to find the field equations of gravitation in the absence of matter. Here we again apply the method

* It is only between the second (and first) derivatives that, by § 12, the relations $B^{\rho}_{\mu\sigma\tau} = 0$ subsist.

employed in the preceding paragraph in formulating the equations of motion of the material point. A special case in which the required equations must in any case be satisfied is that of the special theory of relativity, in which the $g_{\mu\nu}$ have certain constant values. Let this be the case in a certain finite space in relation to a definite system of co-ordinates K_0. Relatively to this system all the components of the Riemann tensor $B^{\rho}_{\mu\sigma\tau}$, defined in (43), vanish. For the space under consideration they then vanish, also in any other system of co-ordinates.

Thus the required equations of the matter-free gravitational field must in any case be satisfied if all $B^{\rho}_{\mu\sigma\tau}$ vanish. But this condition goes too far. For it is clear that, e.g., the gravitational field generated by a material point in its environment certainly cannot be "transformed away" by any choice of the system of co-ordinates, i.e. it cannot be transformed to the case of constant $g_{\mu\nu}$.

This prompts us to require for the matter-free gravitational field that the symmetrical tensor $G_{\mu\nu}$, derived from the tensor $B^{\rho}_{\mu\nu\tau}$, shall vanish. Thus we obtain ten equations for the ten quantities $g_{\mu\nu}$, which are satisfied in the special case of the vanishing of all $B^{\rho}_{\mu\nu\tau}$. With the choice which we have made of a system of co-ordinates, and taking (44) into consideration, the equations for the matter-free field are

$$\left.\begin{aligned} \frac{\partial \Gamma^{\alpha}_{\mu\nu}}{\partial x_{\alpha}} + \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} &= 0 \\ \sqrt{-g} &= 1 \end{aligned}\right\} \quad . \quad . \quad . \quad (47)$$

It must be pointed out that there is only a minimum of arbitrariness in the choice of these equations. For besides $G_{\mu\nu}$ there is no tensor of second rank which is formed from the $g_{\mu\nu}$ and its derivatives, contains no derivations higher than second, and is linear in these derivatives.*

These equations, which proceed, by the method of pure

* Properly speaking, this can be affirmed only of the tensor

$$G_{\mu\nu} + \lambda g_{\mu\nu} g^{\alpha\beta} G_{\alpha\beta},$$

where λ is a constant. If, however, we set this tensor $= 0$, we come back again to the equations $G_{\mu\nu} = 0$.

mathematics, from the requirement of the general theory of relativity, give us, in combination with the equations of motion (46), to a first approximation Newton's law of attraction, and to a second approximation the explanation of the motion of the perihelion of the planet Mercury discovered by Leverrier (as it remains after corrections for perturbation have been made). These facts must, in my opinion, be taken as a convincing proof of the correctness of the theory.

§ 15. The Hamiltonian Function for the Gravitational Field. Laws of Momentum and Energy

To show that the field equations correspond to the laws of momentum and energy, it is most convenient to write them in the following Hamiltonian form :—

$$\left.\begin{aligned} \delta\int \mathrm{H}d\tau &= 0 \\ \mathrm{H} &= g^{\mu\nu}\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} \\ \sqrt{-g} &= 1 \end{aligned}\right\} \quad . \quad . \quad . \quad (47a)$$

where, on the boundary of the finite four-dimensional region of integration which we have in view, the variations vanish.

We first have to show that the form (47a) is equivalent to the equations (47). For this purpose we regard H as a function of the $g^{\mu\nu}$ and the $g^{\mu\nu}_{\sigma}$ $(= \partial g^{\mu\nu}/\partial x_\sigma)$.

Then in the first place

$$\begin{aligned} \delta\mathrm{H} &= \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}\,\delta g^{\mu\nu} + 2g^{\mu\nu}\Gamma^{\alpha}_{\mu\beta}\,\delta\Gamma^{\beta}_{\nu\alpha} \\ &= -\,\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}\,\delta g^{\mu\nu} + 2\Gamma^{\alpha}_{\mu\beta}\,\delta(g^{\mu\nu}\Gamma^{\beta}_{\nu\alpha}). \end{aligned}$$

But

$$\delta\left(g^{\mu\nu}\Gamma^{\beta}_{\nu\alpha}\right) = -\tfrac{1}{2}\delta\left[g^{\mu\nu}g^{\beta\lambda}\left(\frac{\partial g_{\nu\lambda}}{\partial x_\alpha} + \frac{\partial g_{\alpha\lambda}}{\partial x_\nu} - \frac{\partial g_{\alpha\nu}}{\partial x_\lambda}\right)\right].$$

The terms arising from the last two terms in round brackets are of different sign, and result from each other (since the denomination of the summation indices is immaterial) through interchange of the indices μ and β. They cancel each other in the expression for $\delta\mathrm{H}$, because they are multiplied by the quantity $\Gamma^{\alpha}_{\mu\beta}$, which is symmetrical with respect to the indices μ and β. Thus there remains only the first term in round brackets to be considered, so that, taking (31) into account, we obtain

$$\delta H = - \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}\delta g^{\mu\nu} + \Gamma^{\alpha}_{\mu\beta}\delta g^{\mu\beta}_{\alpha}.$$

Thus

$$\left.\begin{aligned} \frac{\partial H}{\partial g^{\mu\nu}} &= - \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} \\ \frac{\partial H}{\partial g^{\mu\nu}_{\sigma}} &= \Gamma^{\sigma}_{\mu\nu} \end{aligned}\right\} \quad . \quad . \quad . \quad (48)$$

Carrying out the variation in (47a), we get in the first place

$$\frac{\partial}{\partial x_{\alpha}}\left(\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial g^{\mu\nu}} = 0, \quad . \quad . \quad . \quad (47b)$$

which, on account of (48), agrees with (47), as was to be proved.

If we multiply (47b) by $g^{\mu\nu}_{\sigma}$, then because

$$\frac{\partial g^{\mu\nu}_{\sigma}}{\partial x_{\alpha}} = \frac{\partial g^{\mu\nu}_{\alpha}}{\partial x_{\sigma}}$$

and, consequently,

$$g^{\mu\nu}_{\sigma}\frac{\partial}{\partial x_{\alpha}}\left(\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) = \frac{\partial}{\partial x_{\alpha}}\left(g^{\mu\nu}_{\sigma}\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\frac{\partial g^{\mu\nu}_{\alpha}}{\partial x_{\sigma}},$$

we obtain the equation

$$\frac{\partial}{\partial x_{\alpha}}\left(g^{\mu\nu}_{\sigma}\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial x_{\sigma}} = 0$$

or *

$$\left.\begin{aligned} \frac{\partial t^{\alpha}_{\sigma}}{\partial x_{\alpha}} &= 0 \\ -2\kappa t^{\alpha}_{\sigma} &= g^{\mu\nu}_{\sigma}\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}} - \delta^{\alpha}_{\sigma}H \end{aligned}\right\} \quad . \quad . \quad . \quad (49)$$

where, on account of (48), the second equation of (47), and (34)

$$\kappa t^{\alpha}_{\sigma} = \tfrac{1}{2}\delta^{\alpha}_{\sigma}g^{\mu\nu}\Gamma^{\lambda}_{\mu\beta}\Gamma^{\beta}_{\nu\lambda} - g^{\mu\nu}\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\sigma} \quad . \quad . \quad (50)$$

* The reason for the introduction of the factor -2κ will be apparent later.

It is to be noticed that t_σ^α is not a tensor; on the other hand (49) applies to all systems of co-ordinates for which $\sqrt{-g} = 1$. This equation expresses the law of conservation of momentum and of energy for the gravitational field. Actually the integration of this equation over a three-dimensional volume V yields the four equations

$$\frac{d}{dx_4}\int t_\sigma^4 dV = \int (lt_\sigma^1 + mt_\sigma^2 + nt_\sigma^3) dS \quad . \quad . \quad (49a)$$

where l, m, n denote the direction-cosines of direction of the inward drawn normal at the element dS of the bounding surface (in the sense of Euclidean geometry). We recognize in this the expression of the laws of conservation in their usual form. The quantities t_σ^α we call the "energy components" of the gravitational field.

I will now give equations (47) in a third form, which is particularly useful for a vivid grasp of our subject. By multiplication of the field equations (47) by $g^{\nu\sigma}$ these are obtained in the "mixed" form. Note that

$$g^{\nu\sigma}\frac{\partial \Gamma_{\mu\nu}^\alpha}{\partial x_\alpha} = \frac{\partial}{\partial x_\alpha}\left(g^{\nu\sigma}\Gamma_{\mu\nu}^\alpha\right) - \frac{\partial g^{\nu\sigma}}{\partial x_\alpha}\Gamma_{\mu\nu}^\alpha,$$

which quantity, by reason of (34), is equal to

$$\frac{\partial}{\partial x_\alpha}\left(g^{\nu\sigma}\Gamma_{\mu\nu}^\alpha\right) - g^{\nu\beta}\Gamma_{\alpha\beta}^\sigma\Gamma_{\mu\nu}^\alpha - g^{\sigma\beta}\Gamma_{\beta\alpha}^\nu\Gamma_{\mu\nu}^\alpha,$$

or (with different symbols for the summation indices)

$$\frac{\partial}{\partial x_\alpha}\left(g^{\sigma\beta}\Gamma_{\mu\beta}^\alpha\right) - g^{\gamma\delta}\Gamma_{\gamma\beta}^\sigma\Gamma_{\delta\mu}^\beta - g^{\nu\sigma}\Gamma_{\mu\beta}^\alpha\Gamma_{\nu\alpha}^\beta.$$

The third term of this expression cancels with the one arising from the second term of the field equations (47); using relation (50), the second term may be written

$$\kappa(t_\mu^\sigma - \tfrac{1}{2}\delta_\mu^\sigma t),$$

where $t = t_\alpha^\alpha$. Thus instead of equations (47) we obtain

$$\left.\begin{aligned}\frac{\partial}{\partial x_\alpha}\left(g^{\sigma\beta}\Gamma_{\mu\beta}^\alpha\right) &= -\kappa(t_\mu^\sigma - \tfrac{1}{2}\delta_\mu^\sigma t)\\ \sqrt{-g} &= 1\end{aligned}\right\} \quad . \quad . \quad (51)$$

§ 16. The General Form of the Field Equations of Gravitation

The field equations for matter-free space formulated in § 15 are to be compared with the field equation

$$\nabla^2\phi = 0$$

of Newton's theory. We require the equation corresponding to Poisson's equation

$$\nabla^2\phi = 4\pi\kappa\rho,$$

where ρ denotes the density of matter.

The special theory of relativity has led to the conclusion that inert mass is nothing more or less than energy, which finds its complete mathematical expression in a symmetrical tensor of second rank, the energy-tensor. Thus in the general theory of relativity we must introduce a corresponding energy-tensor of matter T^{α}_{σ}, which, like the energy-components t_{σ} [equations (49) and (50)] of the gravitational field, will have mixed character, but will pertain to a symmetrical covariant tensor.*

The system of equation (51) shows how this energy-tensor (corresponding to the density ρ in Poisson's equation) is to be introduced into the field equations of gravitation. For if we consider a complete system (e.g. the solar system), the total mass of the system, and therefore its total gravitating action as well, will depend on the total energy of the system, and therefore on the ponderable energy together with the gravitational energy. This will allow itself to be expressed by introducing into (51), in place of the energy-components of the gravitational field alone, the sums $t^{\sigma}_{\mu} + T^{\sigma}_{\mu}$ of the energy-components of matter and of gravitational field. Thus instead of (51) we obtain the tensor equation

$$\left.\begin{aligned} \frac{\partial}{\partial x_{\alpha}}(g^{\sigma\beta}T^{\alpha}_{\mu\beta}) &= -\kappa[(t^{\sigma}_{\mu} + T^{\sigma}_{\mu}) - \tfrac{1}{2}\delta^{\sigma}_{\mu}(t + T)], \\ \sqrt{-g} &= 1 \end{aligned}\right\} \quad . \quad (52)$$

where we have set $T = T^{\mu}_{\mu}$ (Laue's scalar). These are the

* $g_{\alpha\tau}T^{\alpha}_{\sigma} = T_{\sigma\tau}$ and $g^{\sigma\beta}T^{\alpha}_{\sigma} = T^{\alpha\beta}$ are to be symmetrical tensors.

required general field equations of gravitation in mixed form. Working back from these, we have in place of (47)

$$\left.\begin{aligned}\frac{\partial}{\partial x_\alpha}\Gamma^\alpha_{\mu\nu} + \Gamma^\alpha_{\mu\beta}\Gamma^\beta_{\nu\alpha} &= -\kappa(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T),\\ \sqrt{-g} &= 1\end{aligned}\right\} \quad . \quad (53)$$

It must be admitted that this introduction of the energy-tensor of matter is not justified by the relativity postulate alone. For this reason we have here deduced it from the requirement that the energy of the gravitational field shall act gravitatively in the same way as any other kind of energy. But the strongest reason for the choice of these equations lies in their consequence, that the equations of conservation of momentum and energy, corresponding exactly to equations (49) and (49a), hold good for the components of the total energy. This will be shown in § 17.

§ 17. The Laws of Conservation in the General Case

Equation (52) may readily be transformed so that the second term on the right-hand side vanishes. Contract (52) with respect to the indices μ and σ, and after multiplying the resulting equation by $\frac{1}{2}\delta^\sigma_\mu$, subtract it from equation (52). This gives

$$\frac{\partial}{\partial x_\alpha}(g^{\sigma\beta}\Gamma^\alpha_{\mu\beta} - \tfrac{1}{2}\delta^\sigma_\mu g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}) = -\kappa(t^\sigma_\mu + T^\sigma_\mu). \quad . \quad (52a)$$

On this equation we perform the operation $\partial/\partial x_\sigma$. We have

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left(g^\sigma\Gamma^{\ \alpha}_{\beta\mu}\right) = -\tfrac{1}{2}\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left[g^{\sigma\beta}g^{\alpha\lambda}\left(\frac{\partial g_{\mu\lambda}}{\partial x_\beta} + \frac{\partial g_{\beta\lambda}}{\partial x_\mu} - \frac{\partial g_{\mu\beta}}{\partial x_\lambda}\right)\right].$$

The first and third terms of the round brackets yield contributions which cancel one another, as may be seen by interchanging, in the contribution of the third term, the summation indices α and σ on the one hand, and β and λ on the other. The second term may be re-modelled by (31), so that we have

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left(g^{\sigma\beta}\Gamma^\alpha_{\mu\beta}\right) = \tfrac{1}{2}\frac{\partial^3 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta \partial x_\mu} \quad . \quad . \quad (54)$$

The second term on the left-hand side of (52a) yields in the first place

$$-\frac{1}{2}\frac{\partial^2}{\partial x_\alpha \partial x_\mu}\left(g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}\right)$$

or

$$\frac{1}{4}\frac{\partial^2}{\partial x_\alpha \partial x_\mu}\left[g^{\lambda\beta}g^{\alpha\delta}\left(\frac{\partial g_{\delta\lambda}}{\partial x_\beta}+\frac{\partial g_{\delta\beta}}{\partial x_\lambda}-\frac{\partial g_{\lambda\beta}}{\partial x_\delta}\right)\right].$$

With the choice of co-ordinates which we have made, the term deriving from the last term in round brackets disappears by reason of (29). The other two may be combined, and together, by (31), they give

$$-\frac{1}{2}\frac{\partial^3 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta \partial x_\mu},$$

so that in consideration of (54), we have the identity

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left(g^{\rho\beta}\Gamma_{\mu\beta}-\frac{1}{2}\delta^\sigma_\mu g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}\right)\equiv 0 \quad . \quad . \quad (55)$$

From (55) and (52a), it follows that

$$\frac{\partial(t^\sigma_\mu+\mathrm{T}^\sigma_\mu)}{\partial x_\sigma}=0. \quad . \quad . \quad . \quad (56)$$

Thus it results from our field equations of gravitation that the laws of conservation of momentum and energy are satisfied. This may be seen most easily from the consideration which leads to equation (49a); except that here, instead of the energy components t^σ of the gravitational field, we have to introduce the totality of the energy components of matter and gravitational field.

§ 18. The Laws of Momentum and Energy for Matter, as a Consequence of the Field Equations

Multiplying (53) by $\partial g^{\mu\nu}/\partial x_\sigma$, we obtain, by the method adopted in § 15, in view of the vanishing of

$$g_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma},$$

the equation

$$\frac{\partial t^\alpha_\sigma}{\partial x_\alpha}+\frac{1}{2}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}\mathrm{T}_{\mu\nu}=0,$$

or, in view of (56),

$$\frac{\partial T_\sigma^\alpha}{\partial x_\alpha} + \frac{1}{2}\frac{\partial g^{\mu\nu}}{\partial x_\sigma} T_{\mu\nu} = 0 \quad . \quad . \quad . \quad (57)$$

Comparison with (41b) shows that with the choice of system of co-ordinates which we have made, this equation predicates nothing more or less than the vanishing of divergence of the material energy-tensor. Physically, the occurrence of the second term on the left-hand side shows that laws of conservation of momentum and energy do not apply in the strict sense for matter alone, or else that they apply only when the $g^{\mu\nu}$ are constant, i.e. when the field intensities of gravitation vanish. This second term is an expression for momentum, and for energy, as transferred per unit of volume and time from the gravitational field to matter. This is brought out still more clearly by re-writing (57) in the sense of (41) as

$$\frac{\partial T_\sigma^\alpha}{\partial x_\alpha} = -\Gamma_{\alpha\sigma}^\beta T_\beta^\alpha \quad . \quad . \quad . \quad (57a)$$

The right side expresses the energetic effect of the gravitational field on matter.

Thus the field equations of gravitation contain four conditions which govern the course of material phenomena. They give the equations of material phenomena completely, if the latter is capable of being characterized by four differential equations independent of one another.*

D. Material Phenomena

The mathematical aids developed in part B enable us forthwith to generalize the physical laws of matter (hydrodynamics, Maxwell's electrodynamics), as they are formulated in the special theory of relativity, so that they will fit in with the general theory of relativity. When this is done, the general principle of relativity does not indeed afford us a further limitation of possibilities ; but it makes us acquainted with the influence of the gravitational field on all processes,

* On this question cf. H. Hilbert, Nachr. d. K. Gesellsch. d. Wiss. zu Gottingen, Math.-phys. Klasse, 1915, p. 3.

without our having to introduce any new hypothesis whatever.

Hence it comes about that it is not necessary to introduce definite assumptions as to the physical nature of matter (in the narrower sense). In particular it may remain an open question whether the theory of the electromagnetic field in conjunction with that of the gravitational field furnishes a sufficient basis for the theory of matter or not. The general postulate of relativity is unable on principle to tell us anything about this. It must remain to be seen, during the working out of the theory, whether electromagnetics and the doctrine of gravitation are able in collaboration to perform what the former by itself is unable to do.

§ 19. Euler's Equations for a Frictionless Adiabatic Fluid

Let p and ρ be two scalars, the former of which we call the "pressure," the latter the "density" of a fluid; and let an equation subsist between them. Let the contravariant symmetrical tensor

$$T^{\alpha\beta} = -g^{\alpha\beta}p + \rho\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} \quad . \quad . \quad . \quad (58)$$

be the contravariant energy-tensor of the fluid. To it belongs the covariant tensor

$$T_{\mu\nu} = -g_{\mu\nu}p + g_{\mu\alpha}g_{\mu\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds}\rho, \quad . \quad . \quad (58a)$$

as well as the mixed tensor *

$$T_\sigma^\alpha = -\delta_\sigma^\alpha p + g_{\sigma\beta}\frac{dx_\beta}{ds}\frac{dx_\alpha}{ds}\rho \quad . \quad . \quad (58b)$$

Inserting the right-hand side of (58b) in (57a), we obtain the Eulerian hydrodynamical equations of the general theory of relativity. They give, in theory, a complete solution of the problem of motion, since the four equations (57a), together

* For an observer using a system of reference in the sense of the special theory of relativity for an infinitely small region, and moving with it, the density of energy T_4^4 equals $\rho - p$. This gives the definition of ρ. Thus ρ is not constant for an incompressible fluid.

with the given equation between p and ρ, and the equation

$$g_{\alpha\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 1,$$

are sufficient, $g_{\alpha\beta}$ being given, to define the six unknowns

$$p, \rho, \frac{dx_1}{ds}, \frac{dx_2}{ds}, \frac{dx_3}{ds}, \frac{dx_4}{ds}.$$

If the $g_{\mu\nu}$ are also unknown, the equations (53) are brought in. These are eleven equations for defining the ten functions $g_{\mu\nu}$, so that these functions appear over-defined. We must remember, however, that the equations (57a) are already contained in the equations (53), so that the latter represent only seven independent equations. There is good reason for this lack of definition, in that the wide freedom of the choice of co-ordinates causes the problem to remain mathematically undefined to such a degree that three of the functions of space may be chosen at will.*

§ 20. Maxwell's Electromagnetic Field Equations for Free Space

Let ϕ_ν be the components of a covariant vector—the electromagnetic potential vector. From them we form, in accordance with (36), the components $F_{\rho\sigma}$ of the covariant six-vector of the electromagnetic field, in accordance with the system of equations

$$F_{\rho\sigma} = \frac{\partial\phi_\rho}{\partial x_\sigma} - \frac{\partial\phi_\sigma}{\partial x_\rho} \quad . \quad . \quad . \quad (59)$$

It follows from (59) that the system of equations

$$\frac{\partial F_{\rho\sigma}}{\partial x_\tau} + \frac{\partial F_{\sigma\tau}}{\partial x_\rho} + \frac{\partial F_{\tau\rho}}{\partial x_\sigma} = 0 \quad . \quad . \quad . \quad (60)$$

is satisfied, its left side being, by (37), an antisymmetrical tensor of the third rank. System (60) thus contains essentially four equations which are written out as follows:—

* On the abandonment of the choice of co-ordinates with $g = -1$, there remain *four* functions of space with liberty of choice, corresponding to the four arbitrary functions at our disposal in the choice of co-ordinates.

$$\left.\begin{aligned}\frac{\partial F_{23}}{\partial x_4}+\frac{\partial F_{34}}{\partial x_2}+\frac{\partial F_{42}}{\partial x_3}&=0\\ \frac{\partial F_{34}}{\partial x_1}+\frac{\partial F_{41}}{\partial x_3}+\frac{\partial F_{13}}{\partial x_4}&=0\\ \frac{\partial F_{41}}{\partial x_2}+\frac{\partial F_{12}}{\partial x_4}+\frac{\partial F_{24}}{\partial x_1}&=0\\ \frac{\partial F_{12}}{\partial x_3}+\frac{\partial F_{23}}{\partial x_1}+\frac{\partial F_{31}}{\partial x_2}&=0\end{aligned}\right\} \qquad (60a)$$

This system corresponds to the second of Maxwell's systems of equations. We recognize this at once by setting

$$\left.\begin{aligned}F_{23}&=H_x, & F_{14}&=E_x\\ F_{31}&=H_y, & F_{24}&=E_y\\ F_{12}&=H_z, & F_{34}&=E_z\end{aligned}\right\} \quad . \quad . \quad . \quad (61)$$

Then in place of (60a) we may set, in the usual notation of three-dimensional vector analysis,

$$\left.\begin{aligned}-\frac{\partial H}{\partial t}&=\text{curl } E\\ \text{div } H&=0\end{aligned}\right\} \quad . \quad . \quad . \quad (60b)$$

We obtain Maxwell's first system by generalizing the form given by Minkowski. We introduce the contravariant six-vector associated with $F^{\alpha\beta}$

$$F^{\mu\nu}=g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta} \quad . \quad . \quad . \quad (62)$$

and also the contravariant vector J^{μ} of the density of the electric current. Then, taking (40) into consideration, the following equations will be invariant for any substitution whose invariant is unity (in agreement with the chosen co-ordinates) :—

$$\frac{\partial}{\partial x_\nu}F^{\mu\nu}=J^{\mu} \quad . \quad . \quad . \quad . \quad (63)$$

Let

$$\left.\begin{aligned}F^{23}&=H'_x, & F^{14}&=-E'_x\\ F^{31}&=H'_y, & F^{24}&=-E'_y\\ F^{12}&=H'_z, & F^{34}&=-E'_z\end{aligned}\right\} \quad . \quad . \quad (64)$$

which quantities are equal to the quantities $H_x \ldots E^z$ in the special case of the restricted theory of relativity ; and in addition

$$J^1 = j_x,\ J^2 = j_y,\ J^3 = j_z,\ J^4 = \rho,$$

we obtain in place of (63)

$$\left.\begin{aligned} \frac{\partial E'}{\partial t} + j &= \operatorname{curl} H' \\ \operatorname{div} E' &= \rho \end{aligned}\right\} \quad . \quad . \quad . \quad (63a)$$

The equations (60), (62), and (63) thus form the generalization of Maxwell's field equations for free space, with the convention which we have established with respect to the choice of co-ordinates.

The Energy-components of the Electromagnetic Field.—We form the inner product

$$\kappa_\sigma = F_{\sigma\mu} J^\mu \quad . \quad . \quad . \quad . \quad (65)$$

By (61) its components, written in the three-dimensional manner, are

$$\left.\begin{aligned} \kappa_1 &= \rho E_x + [j \,.\, H]^x \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ \kappa_4 &= -(jE) \end{aligned}\right\} \quad . \quad . \quad . \quad (65a)$$

κ_σ is a covariant vector the components of which are equal to the negative momentum, or, respectively, the energy, which is transferred from the electric masses to the electromagnetic field per unit of time and volume. If the electric masses are free, that is, under the sole influence of the electromagnetic field, the covariant vector κ_σ will vanish.

To obtain the energy-components T_σ^ν of the electromagnetic field, we need only give to equation $\kappa_\sigma = 0$ the form of equation (57). From (63) and (65) we have in the first place

$$\kappa_\sigma = F_{\sigma\mu}\frac{\partial F^{\mu\nu}}{\partial x_\nu} = \frac{\partial}{\partial x_\nu}(F_{\sigma\mu}F^{\mu\nu}) - F^{\mu\nu}\frac{\partial F_{\sigma\mu}}{\partial x_\nu}.$$

The second term of the right-hand side, by reason of (60), permits the transformation

$$F^{\mu\nu}\frac{\partial F_{\sigma\mu}}{\partial x_\nu} = -\tfrac{1}{2}F^{\mu\nu}\frac{\partial F_{\mu\nu}}{\partial x_\sigma} = -\tfrac{1}{2}g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}\frac{\partial F_{\mu\nu}}{\partial x_\sigma},$$

which latter expression may, for reasons of symmetry, also be written in the form

$$-\tfrac{1}{4}\left[g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}\frac{\partial F_{\mu\nu}}{\partial x_\sigma} + g^{\mu\alpha}g^{\nu\beta}\frac{\partial F_{\alpha\beta}}{\partial x_\sigma}F_{\mu\nu}\right].$$

But for this we may set

$$-\tfrac{1}{4}\frac{\partial}{\partial x_\sigma}(g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}F_{\mu\nu}) + \tfrac{1}{4}F_{\alpha\beta}F_{\mu\nu}\frac{\partial}{\partial x_\sigma}(g^{\mu\alpha}g^{\nu\beta}).$$

The first of these terms is written more briefly

$$-\tfrac{1}{4}\frac{\partial}{\partial x_\sigma}(F^{\mu\nu}F_{\mu\nu});$$

the second, after the differentiation is carried out, and after some reduction, results in

$$-\tfrac{1}{2}F^{\mu\tau}F_{\mu\nu}g^{\nu\rho}\frac{\partial g_{\sigma\tau}}{\partial x_\sigma}.$$

Taking all three terms together we obtain the relation

$$\kappa_\sigma = \frac{\partial T^\nu_\sigma}{\partial x_\nu} - \tfrac{1}{2}g^{\tau\mu}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}T^\nu_\tau \quad . \quad . \quad . \quad (66)$$

where

$$T^\nu_\sigma = -F_{\sigma\alpha}F^{\nu\alpha} + \tfrac{1}{4}\delta^\nu_\sigma F_{\alpha\beta}F^{\alpha\beta}.$$

Equation (66), if κ_σ vanishes, is, on account of (30), equivalent to (57) or (57*a*) respectively. Therefore the T^ν_σ are the energy-components of the electromagnetic field. With the help of (61) and (64), it is easy to show that these energy-components of the electromagnetic field in the case of the special theory of relativity give the well-known Maxwell-Poynting expressions.

We have now deduced the general laws which are satisfied by the gravitational field and matter, by consistently using a system of co-ordinates for which $\sqrt{-g} = 1$. We have thereby achieved a considerable simplification of formulæ and calculations, without failing to comply with the requirement of general covariance; for we have drawn our equations from generally covariant equations by specializing the system of co-ordinates.

Still the question is not without a formal interest, whether with a correspondingly generalized definition of the energy-components of gravitational field and matter, even without specializing the system of co-ordinates, it is possible to formulate laws of conservation in the form of equation (56), and field equations of gravitation of the same nature as (52) or (52a), in such a manner that on the left we have a divergence (in the ordinary sense), and on the right the sum of the energy-components of matter and gravitation. I have found that in both cases this is actually so. But I do not think that the communication of my somewhat extensive reflexions on this subject would be worth while, because after all they do not give us anything that is materially new.

E

§ 21. Newton's Theory as a First Approximation

As has already been mentioned more than once, the special theory of relativity as a special case of the general theory is characterized by the $g_{\mu\nu}$ having the constant values (4). From what has already been said, this means complete neglect of the effects of gravitation. We arrive at a closer approximation to reality by considering the case where the $g_{\mu\nu}$ differ from the values of (4) by quantities which are small compared with 1, and neglecting small quantities of second and higher order. (First point of view of approximation.)

It is further to be assumed that in the space-time territory under consideration the $g_{\mu\nu}$ at spatial infinity, with a suitable choice of co-ordinates, tend toward the values (4); i.e. we are considering gravitational fields which may be regarded as generated exclusively by matter in the finite region.

It might be thought that these approximations must lead us to Newton's theory. But to that end we still need to approximate the fundamental equations from a second point of view. We give our attention to the motion of a material point in accordance with the equations (16). In the case of the special theory of relativity the components

$$\frac{dx_1}{ds}, \frac{dx_2}{ds}, \frac{dx_3}{ds}$$

may take on any values. This signifies that any velocity

$$v = \sqrt{\left(\frac{dx_1}{dx_4}\right)^2 + \left(\frac{dx_2}{dx_4}\right)^2 + \left(\frac{dx_3}{dx_4}\right)^2}$$

may occur, which is less than the velocity of light *in vacuo.* If we restrict ourselves to the case which almost exclusively offers itself to our experience, of v being small as compared with the velocity of light, this denotes that the components

$$\frac{dx_1}{ds}, \frac{dx_2}{ds}, \frac{dx_3}{ds}$$

are to be treated as small quantities, while dx_4/ds, to the second order of small quantities, is equal to one. (Second point of view of approximation.)

Now we remark that from the first point of view of approximation the magnitudes $\Gamma^{\tau}_{\mu\nu}$ are all small magnitudes of at least the first order. A glance at (46) thus shows that in this equation, from the second point of view of approximation, we have to consider only terms for which $\mu = \nu = 4$. Restricting ourselves to terms of lowest order we first obtain in place of (46) the equations

$$\frac{d^2x_\tau}{dt^2} = \Gamma^{\tau}_{44}$$

where we have set $ds = dx_4 = dt$; or with restriction to terms which from the first point of view of approximation are of first order :—

$$\frac{d^2x_\tau}{dt^2} = [44, \tau] \quad (\tau = 1, 2, 3)$$

$$\frac{d^2x_4}{dt^2} = -[44, 4].$$

If in addition we suppose the gravitational field to be a quasi-static field, by confining ourselves to the case where the motion of the matter generating the gravitational field is but slow (in comparison with the velocity of the propagation of light), we may neglect on the right-hand side differentiations with respect to the time in comparison with those with respect to the space co-ordinates, so that we have

$$\frac{d^2x_\tau}{dt^2} = -\tfrac{1}{2}\frac{\partial g_{44}}{\partial x_\tau} \quad (\tau = 1, 2, 3) \quad . \quad . \quad (67)$$

This is the equation of motion of the material point according to Newton's theory, in which $\frac{1}{2}g_{44}$ plays the part of the gravitational potential. What is remarkable in this result is that the component g_{44} of the fundamental tensor alone defines, to a first approximation, the motion of the material point.

We now turn to the field equations (53). Here we have to take into consideration that the energy-tensor of "matter" is almost exclusively defined by the density of matter in the narrower sense, i.e. by the second term of the right-hand side of (58) [or, respectively, (58a) or (58b)]. If we form the approximation in question, all the components vanish with the one exception of $T_{44} = \rho = T$. On the left-hand side of (53) the second term is a small quantity of second order; the first yields, to the approximation in question,

$$\frac{\partial}{\partial x_1}[\mu\nu, 1] + \frac{\partial}{\partial x_2}[\mu\nu, 2] + \frac{\partial}{\partial x_3}[\mu\nu, 3] - \frac{\partial}{\partial x_4}[\mu\nu, 4].$$

For $\mu = \nu = 4$, this gives, with the omission of terms differentiated with respect to time,

$$-\tfrac{1}{2}\left(\frac{\partial^2 g_{44}}{\partial x_1^2} + \frac{\partial^2 g_{44}}{\partial x_2^2} + \frac{\partial^2 g_{44}}{\partial x_3^2}\right) = -\tfrac{1}{2}\nabla^2 g_{44}.$$

The last of equations (53) thus yields

$$\nabla^2 g_{44} = \kappa\rho \quad . \quad . \quad . \quad . \quad . \quad (68)$$

The equations (67) and (68) together are equivalent to Newton's law of gravitation.

By (67) and (68) the expression for the gravitational potential becomes

$$-\frac{\kappa}{8\pi}\int\frac{\rho d\tau}{r} \quad . \quad . \quad . \quad . \quad . \quad (68a)$$

while Newton's theory, with the unit of time which we have chosen, gives

$$-\frac{K}{c^2}\int\frac{\rho d\tau}{r}$$

in which K denotes the constant $6 \cdot 7 \times 10^{-8}$, usually called the constant of gravitation. By comparison we obtain

$$\kappa = \frac{8\pi K}{c^2} = 1 \cdot 87 \times 10^{-27} \quad . \quad . \quad (69)$$

§ 22. Behaviour of Rods and Clocks in the Static Gravitational Field. Bending of Light-rays. Motion of the Perihelion of a Planetary Orbit

To arrive at Newton's theory as a first approximation we had to calculate only one component, g_{44}, of the ten $g_{\mu\nu}$ of the gravitational field, since this component alone enters into the first approximation, (67), of the equation for the motion of the material point in the gravitational field. From this, however, it is already apparent that other components of the $g_{\mu\nu}$ must differ from the values given in (4) by small quantities of the first order. This is required by the condition $g = -1$.

For a field-producing point mass at the origin of co-ordinates, we obtain, to the first approximation, the radially symmetrical solution

$$\left.\begin{aligned} g_{\rho\sigma} &= -\delta_{\rho\sigma} - a\frac{x_\rho x_\sigma}{r^3} \; (\rho, \sigma = 1, 2, 3) \\ g_{\rho 4} &= g_{4\rho} = 0 \qquad\quad (\rho = 1, 2, 3) \\ g_{44} &= 1 - \frac{a}{r} \end{aligned}\right\} \quad . \quad (70)$$

where $\delta_{\rho\sigma}$ is 1 or 0, respectively, accordingly as $\rho = \sigma$ or $\rho \neq \sigma$, and r is the quantity $+\sqrt{x_1^2 + x_2^2 + x_3^2}$. On account of (68a)

$$a = \frac{\kappa M}{4\pi}, \quad . \quad . \quad . \quad . \quad (70a)$$

if M denotes the field-producing mass. It is easy to verify that the field equations (outside the mass) are satisfied to the first order of small quantities.

We now examine the influence exerted by the field of the mass M upon the metrical properties of space. The relation

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu.$$

always holds between the "locally" (§ 4) measured lengths and times ds on the one hand, and the differences of co-ordinates dx_ν on the other hand.

For a unit-measure of length laid "parallel" to the axis of x, for example, we should have to set $ds^2 = -1$; $dx_2 = dx_3 = dx_4 = 0$. Therefore $-1 = g_{11}dx_1^2$. If, in addition, the unit-measure lies on the axis of x, the first of equations (70) gives

$$g_{11} = -\left(1 + \frac{a}{r}\right).$$

From these two relations it follows that, correct to a first order of small quantities,

$$dx = 1 - \frac{a}{2r} \quad . \quad . \quad . \quad . \quad (71)$$

The unit measuring-rod thus appears a little shortened in relation to the system of co-ordinates by the presence of the gravitational field, if the rod is laid along a radius.

In an analogous manner we obtain the length of co-ordinates in tangential direction if, for example, we set

$ds^2 = -1$; $dx_1 = dx_3 = dx_4 = 0$; $x_1 = r$, $x_2 = x_3 = 0$.
The result is

$$-1 = g_{22}dx_2^2 = -dx_2^2 \quad . \quad . \quad . \quad (71a)$$

With the tangential position, therefore, the gravitational field of the point of màss has no influence on the length of a rod.

Thus Euclidean geometry does not hold even to a first approximation in the gravitational field, if we wish to take one and the same rod, independently of its place and orientation, as a realization of the same interval; although, to be sure, a glance at (70a) and (69) shows that the deviations to be expected are much too slight to be noticeable in measurements of the earth's surface.

Further, let us examine the rate of a unit clock, which is arranged to be at rest in a static gravitational field. Here we have for a clock period $ds = 1$; $dx_1 = dx_2 = dx_3 = 0$
Therefore

$$1 = g_{44}dx_4^2;$$

$$dx_4 = \frac{1}{\sqrt{g_{44}}} = \frac{1}{\sqrt{(1 + (g_{44} - 1))}} = 1 - \tfrac{1}{2}(g_{44} - 1)$$

or

$$dx_4 = 1 + \frac{\kappa}{8\pi}\int \rho\frac{d\tau}{r} \quad . \quad . \quad . \quad (72)$$

Thus the clock goes more slowly if set up in the neighbourhood of ponderable masses. From this it follows that the spectral lines of light reaching us from the surface of large stars must appear displaced towards the red end of the spectrum.*

We now examine the course of light-rays in the static gravitational field. By the special theory of relativity the velocity of light is given by the equation

$$- dx_1^2 - dx_2 - dx_3^2 + dx_4^2 = 0$$

and therefore by the general theory of relativity by the equation

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu = 0 \quad . \quad . \quad . \quad (73)$$

If the direction, i.e. the ratio $dx_1 : dx_2 : dx_3$ is given, equation (73) gives the quantities

$$\frac{dx_1}{dx_4}, \frac{dx_2}{dx_4}, \frac{dx_3}{dx_4}$$

and accordingly the velocity

$$\sqrt{\left(\frac{dx_1}{dx_4}\right)^2 + \left(\frac{dx_2}{dx_4}\right)^2 + \left(\frac{dx_3}{dx_4}\right)^2} = \gamma$$

defined in the sense of Euclidean geometry. We easily recognize that the course of the light-rays must be bent with regard to the system of co-ordinates, if the $g_{\mu\nu}$ are not constant. If n is a direction perpendicular to the propagation of light, the Huyghens principle shows that the light-ray, envisaged in the plane (γ, n), has the curvature $-\partial\gamma/\partial n$.

We examine the curvature undergone by a ray of light passing by a mass M at the distance Δ. If we choose the system of co-ordinates in agreement with the accompanying diagram, the total bending of the ray (calculated positively if

* According to E. Freundlich, spectroscopical observations on fixed stars of certain types indicate the existence of an effect of this kind, but a crucial test of this consequence has not yet been made.

concave towards the origin) is given in sufficient approximation by

$$B = \int_{-\infty}^{+\infty} \frac{\partial \gamma}{\partial x_1} dx_2,$$

while (73) and (70) give

$$\gamma = \sqrt{\left(-\frac{g_{44}}{g_{22}}\right)} = 1 - \frac{a}{2r}\left(1 + \frac{x_2^2}{r^2}\right).$$

Carrying out the calculation, this gives

$$B = \frac{2a}{\Delta} = \frac{\kappa M}{2\pi\Delta}. \qquad . \quad . \quad . \quad (74)$$

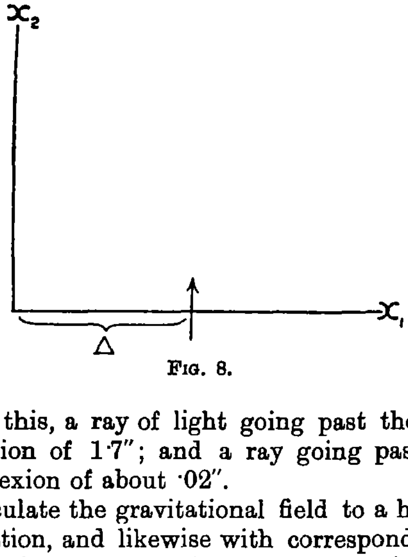


FIG. 8.

According to this, a ray of light going past the sun undergoes a deflexion of 1·7″; and a ray going past the planet Jupiter a deflexion of about ·02″.

If we calculate the gravitational field to a higher degree of approximation, and likewise with corresponding accuracy the orbital motion of a material point of relatively infinitely small mass, we find a deviation of the following kind from the Kepler-Newton laws of planetary motion. The orbital ellipse of a planet undergoes a slow rotation, in the direction of motion, of amount

$$\epsilon = 24\pi^3 \frac{a^2}{T^2 c^2 (1 - e^2)} \qquad . \quad . \quad . \quad (75)$$

per revolution. In this formula *a* denotes the major semi-axis, *c* the velocity of light in the usual measurement, *e* the eccentricity, T the time of revolution in seconds.*

Calculation gives for the planet Mercury a rotation of the orbit of 43″ per century, corresponding exactly to astronomical observation (Leverrier); for the astronomers have discovered in the motion of the perihelion of this planet, after allowing for disturbances by other planets, an inexplicable remainder of this magnitude.

* For the calculation I refer to the original papers: A. Einstein, Sitzungsber. d. Preuss. Akad. d. Wiss., 1915, p. 831; K. Schwarzschild, *ibid.*, 1916, p. 189.

HAMILTON'S PRINCIPLE AND THE GENERAL THEORY OF RELATIVITY

BY

A. EINSTEIN

Translated from "Hamiltonsches Princip und allgemeine Relativitätstheorie," Sitzungsberichte der Preussischen Akad. d. Wissenschaften, 1916.

HAMILTON'S PRINCIPLE AND THE GENERAL THEORY OF RELATIVITY

By A. EINSTEIN

THE general theory of relativity has recently been given in a particularly clear form by H. A. Lorentz and D. Hilbert,* who have deduced its equations from one single principle of variation. The same thing will be done in the present paper. But my purpose here is to present the fundamental connexions in as perspicuous a manner as possible, and in as general terms as is permissible from the point of view of the general theory of relativity. In particular we shall make as few specializing assumptions as possible, in marked contrast to Hilbert's treatment of the subject. On the other hand, in antithesis to my own most recent treatment of the subject, there is to be complete liberty in the choice of the system of co-ordinates.

§ 1. The Principle of Variation and the Field-equations of Gravitation and Matter

Let the gravitational field be described as usual by the tensor † of the $g_{\mu\nu}$ (or the $g^{\mu\nu}$); and matter, including the electromagnetic field, by any number of space-time functions $q_{(\rho)}$. How these functions may be characterized in the theory of invariants does not concern us. Further, let $\mathfrak{H}$ be a function of the

$$g^{\mu\nu}, g^{\mu\nu}_{\sigma}\left(=\frac{\partial g^{\mu\nu}}{\partial x_{\sigma}}\right) \text{ and } g^{\mu\nu}_{\sigma\tau}\left(=\frac{\partial^2 g^{\mu\nu}}{\partial x_{\sigma}\partial x_{\tau}}\right), \text{the } q_{(\rho)} \text{ and } q_{(\rho)\alpha}\left(=\frac{\partial q_{(\rho)}}{\partial x_{\alpha}}\right).$$

* Four papers by Lorentz in the Publications of the Koninkl. Akad. van Wetensch. te Amsterdam, 1915 and 1916; D. Hilbert, Gottinger Nachr., 1915, Part 3.

† No use is made for the present of the tensor character of the $g_{\mu\nu}$.

The principle of variation

$$\delta\int \mathfrak{H} d\tau = 0 \quad . \quad . \quad . \quad . \quad (1)$$

then gives us as many differential equations as there are functions $g_{\mu\nu}$ and $q_{(\rho)}$ to be defined, if the $g^{\mu\nu}$ and $q_{(\rho)}$ are varied independently of one another, and in such a way that at the limits of integration the $\delta q_{(\rho)}$, $\delta g^{\mu\nu}$, and $\frac{\partial}{\partial x_\sigma}(\delta g_{\mu\nu})$ all vanish.

We will now assume that $\mathfrak{H}$ is linear in the $g_{\sigma\tau}$, and that the coefficients of the $g^{\mu\nu}_{\sigma\tau}$ depend only on the $g^{\mu\nu}$. We may then replace the principle of variation (1) by one which is more convenient for us. For by appropriate partial integration we obtain

$$\int \mathfrak{H} d\tau = \int \mathfrak{H}^* d\tau + \mathrm{F} \quad . \quad . \quad . \quad (2)$$

where F denotes an integral over the boundary of the domain in question, and $\mathfrak{H}^*$ depends only on the $g^{\mu\nu}$, $g^{\mu\nu}_{\sigma}$, $q_{(\rho)}$, $q_{(\rho)\alpha}$, and no longer on the $g^{\mu\nu}_{\sigma\tau}$. From (2) we obtain, for such variations as are of interest to us,

$$\delta\int \mathfrak{H} d\tau = \delta\int \mathfrak{H}^* d\tau, \quad . \quad . \quad . \quad (3)$$

so that we may replace our principle of variation (1) by the more convenient form

$$\delta\int \mathfrak{H}^* d\tau = 0. \quad . \quad . \quad . \quad (1a)$$

By carrying out the variation of the $g^{\mu\nu}$ and the $q_{(\rho)}$ we obtain, as field-equations of gravitation and matter, the equations †

$$\frac{\partial}{\partial x_\alpha}\left(\frac{\partial \mathfrak{H}^*}{\partial g^{\mu\nu}_\alpha}\right) - \frac{\partial \mathfrak{H}^*}{\partial g^{\mu\nu}} = 0 \quad . \quad . \quad . \quad (4)$$

$$\frac{\partial}{\partial x_\alpha}\left(\frac{\partial \mathfrak{H}^*}{\partial q_{(\rho)\alpha}}\right) - \frac{\partial \mathfrak{H}^*}{\partial q_{(\rho)}} = 0 \quad . \quad . \quad . \quad (5)$$

† For brevity the summation symbols are omitted in the formulæ. Indices occurring twice in a term are always to be taken as summed. Thus in (4), for example, $\frac{\partial}{\partial x_\alpha}\left(\frac{\partial \mathfrak{H}^*}{\partial g^{\mu\nu}_\alpha}\right)$ denotes the term $\sum_\alpha \frac{\partial}{\partial x_\alpha}\left(\frac{\partial \mathfrak{H}^*}{\partial g^{\mu\nu}_\alpha}\right)$.

§ 2. Separate Existence of the Gravitational Field

If we make no restrictive assumption as to the manner in which $\mathfrak{H}$ depends on the $g^{\mu\nu}$, $g^{\mu\nu}_{\sigma}$, $g^{\mu\nu}_{\sigma\tau}$, $q_{(\rho)}$, $q_{(\rho)\alpha}$, the energy-components cannot be divided into two parts, one belonging to the gravitational field, the other to matter. To ensure this feature of the theory, we make the following assumption

$$\mathfrak{H} = \mathfrak{G} + \mathfrak{M} \quad . \quad . \quad . \quad . \quad (6)$$

where $\mathfrak{G}$ is to depend only on the $g^{\mu\nu}$, $g^{\mu\nu}_{\sigma}$, $g^{\mu\nu}_{\sigma\tau}$, and $\mathfrak{M}$ only on $g^{\mu\nu}$, $q_{(\rho)}$, $q_{(\rho)\alpha}$. Equations (4), (4a) then assume the form

$$\frac{\partial}{\partial x_\alpha}\left(\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}_\alpha}\right) - \frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}} = \frac{\partial \mathfrak{M}}{\partial g^{\mu\nu}} \quad . \quad . \quad . \quad (7)$$

$$\frac{\partial}{\partial x_\alpha}\left(\frac{\partial \mathfrak{M}}{\partial q_{(\rho)\alpha}}\right) - \frac{\partial \mathfrak{M}}{\partial q_{(\rho)}} = 0 \quad . \quad . \quad . \quad (8)$$

Here $\mathfrak{G}^*$ stands in the same relation to $\mathfrak{G}$ as $\mathfrak{H}^*$ to $\mathfrak{H}$.

It is to be noted carefully that equations (8) or (5) would have to give way to others, if we were to assume $\mathfrak{M}$ or $\mathfrak{H}$ to be also dependent on derivatives of the $q_{(\rho)}$ of order higher than the first. Likewise it might be imaginable that the $q_{(\rho)}$ would have to be taken, not as independent of one another, but as connected by conditional equations. All this is of no importance for the following developments, as these are based solely on the equations (7), which have been found by varying our integral with respect to the $g^{\mu\nu}$.

§ 3. Properties of the Field Equations of Gravitation Conditioned by the Theory of Invariants

We now introduce the assumption that

$$ds^2 = g_{\mu\nu} dx_\mu dx_\nu \quad . \quad . \quad . \quad . \quad (9)$$

is an invariant. This determines the transformational character of the $g_{\mu\nu}$. As to the transformational character of the $q_{(\rho)}$, which describe matter, we make no supposition. On the other hand, let the functions $\mathrm{H} = \frac{\mathfrak{H}}{\sqrt{-g}}$, as well as

$G = \frac{\mathfrak{G}}{\sqrt{-g}}$, and $M = \frac{\mathfrak{M}}{\sqrt{-g}}$, be invariants in relation to any substitutions of space-time co-ordinates. From these assumptions follows the general covariance of the equations (7) and (8), deduced from (1). It further follows that G (apart from a constant factor) must be equal to the scalar of Riemann's tensor of curvature; because there is no other invariant with the properties required for G.† Thereby $\mathfrak{G}^*$ is also perfectly determined, and consequently the left-hand side of field equation (7) as well.‡

From the general postulate of relativity there follow certain properties of the function $\mathfrak{G}^*$ which we shall now deduce. For this purpose we carry through an infinitesimal transformation of the co-ordinates, by setting

$$x'_\nu = x_\nu + \Delta x_\nu \quad . \quad . \quad . \quad . \quad (10)$$

where the Δx_ν are arbitrary, infinitely small functions of the co-ordinates, and x'_ν are the co-ordinates, in the new system, of the world-point having the co-ordinates x_ν in the original system. As for the co-ordinates, so too for any other magnitude ψ, a law of transformation holds good, of the type

$$\psi' = \psi + \Delta\psi,$$

where $\Delta\psi$ must always be expressible by the Δx_ν. From the covariant property of the $g^{\mu\nu}$ we easily deduce for the $g^{\mu\nu}$ and $g^{\mu\nu}_\sigma$ the laws of transformation

$$\Delta g^{\mu\nu} = g^{\mu\alpha}\frac{\partial(\Delta x_\nu)}{\partial x_\alpha} + g^{\nu\alpha}\frac{\partial(\Delta x_\mu)}{\partial x_\alpha} \quad . \quad . \quad (11)$$

$$\Delta g^{\mu\nu}_\sigma = \frac{\partial(\Delta g^{\mu\nu})}{\partial x_\sigma} - g^{\mu\nu}_\alpha\frac{\partial(\Delta x_\alpha)}{\partial x_\sigma} \quad . \quad . \quad . \quad (12)$$

Since $\mathfrak{G}^*$ depends only on the $g^{\mu\nu}$ and $g^{\mu\nu}_\sigma$, it is possible, with the help of (11) and (12), to calculate $\Delta\mathfrak{G}^*$. We thus obtain the equation

$$\sqrt{-g}\Delta\left(\frac{\mathfrak{G}^*}{\sqrt{-g}}\right) = S^\nu_\sigma\frac{\partial(\Delta x_\sigma)}{\partial x_\nu} + 2\frac{\partial\mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}g^{\mu\nu}\frac{\partial^2\Delta x_\sigma}{\partial x_\nu \partial x_\alpha}, \quad (13)$$

† Herein is to be found the reason why the general postulate of relativity leads to a very definite theory of gravitation.

‡ By performing partial integration we obtain

$$\mathfrak{G}^* = \sqrt{-g}\, g^{\mu\nu}[\{\mu\alpha, \beta\}\{\nu\beta, \alpha\} - \{\mu\nu, \alpha\}\{\alpha\beta, \beta\}].$$

where for brevity we have set

$$S_\sigma^\nu = 2\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\sigma}}g^{\mu\nu} + 2\frac{\partial \mathfrak{G}^*}{\partial g_\alpha^{\mu\sigma}}g_\alpha^{\mu\nu} + \mathfrak{G}^*\delta_\sigma^\nu - \frac{\partial \mathfrak{G}^*}{\partial g_\nu^{\mu\alpha}}g_\sigma^{\mu\alpha}. \quad (14)$$

From these two equations we draw two inferences which are important for what follows. We know that $\frac{\mathfrak{G}}{\sqrt{-g}}$ is an invariant with respect to any substitution, but we do not know this of $\frac{\mathfrak{G}^*}{\sqrt{-g}}$. It is easy to demonstrate, however, that the latter quantity is an invariant with respect to any *linear* substitutions of the co-ordinates. Hence it follows that the right side of (13) must always vanish if all $\frac{\partial^2 \Delta x_\sigma}{\partial x_\nu \partial x_\alpha}$ vanish. Consequently $\mathfrak{G}^*$ must satisfy the identity

$$S_\sigma^\nu \equiv 0 \quad . \quad . \quad . \quad . \quad (15)$$

If, further, we choose the Δx_ν so that they differ from zero only in the interior of a given domain, but in infinitesimal proximity to the boundary they vanish, then, with the transformation in question, the value of the boundary integral occurring in equation (2) does not change. Therefore $\Delta F = 0$, and, in consequence,†

$$\Delta\int \mathfrak{G} d\tau = \Delta\int \mathfrak{G}^* d\tau.$$

But the left-hand side of the equation must vanish, since both $\frac{\mathfrak{G}}{\sqrt{-g}}$ and $\sqrt{-g}\, d\tau$ are invariants. Consequently the right-hand side also vanishes. Thus, taking (14), (15), and (16) into consideration, we obtain, in the first place, the equation

$$\int \frac{\partial \mathfrak{G}^*}{\partial g_\alpha^{\mu\sigma}} g^{\mu\nu} \frac{\partial^2(\Delta x_\sigma)}{\partial x_\nu \partial x_\alpha} d\tau = 0 \quad . \quad . \quad . \quad (16)$$

Transforming this equation by two partial integrations, and having regard to the liberty of choice of the Δx_σ, we obtain

† By the introduction of the quantities $\mathfrak{G}$ and $\mathfrak{G}^*$ instead of $\mathfrak{H}$ and $\mathfrak{H}^*$.

the identity

$$\frac{\partial^2}{\partial x_\nu \partial x_\alpha}\left(g^{\mu\nu}\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}\right) \equiv 0 \quad . \quad . \quad . \quad (17)$$

From the two identities (16) and (17), which result from the invariance of $\frac{\mathfrak{G}}{\sqrt{-g}}$, and therefore from the postulate of general relativity, we now have to draw conclusions.

We first transform the field equations (7) of gravitation by mixed multiplication by $g^{\mu\sigma}$. We then obtain (by interchanging the indices σ and ν), as equivalents of the field equations (7), the equations

$$\frac{\partial}{\partial x_\alpha}\left(g^{\mu\nu}\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}\right) = -(\mathfrak{T}^\nu_\sigma + \mathfrak{t}^\nu_\sigma) \quad . \quad . \quad (18)$$

where we have set

$$\mathfrak{T}^\nu_\sigma = -\frac{\partial \mathfrak{M}}{\partial g^{\mu\sigma}}g^{\mu\nu} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$\mathfrak{t}^\nu_\sigma = -\left(\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}g^{\mu\nu}_\alpha + \frac{\partial \mathfrak{G}^*}{\partial g^{\mu\sigma}}g^{\mu\nu}\right) = \tfrac{1}{2}\left(\mathfrak{G}^*\delta^\nu_\sigma - \frac{\partial \mathfrak{G}^*}{\partial g^{\mu\alpha}_\nu}g^{\mu\alpha}_\sigma\right) \quad (20)$$

The last expression for $\mathfrak{t}^\nu_\sigma$ is vindicated by (14) and (15). By differentiation of (18) with respect to x_ν, and summation for ν, there follows, in view of (17),

$$\frac{\partial}{\partial x_\nu}(\mathfrak{T}^\nu_\sigma + \mathfrak{t}^\nu_\sigma) = 0 \quad . \quad . \quad . \quad (21)$$

Equation (21) expresses the conservation of momentum and energy. We call $\mathfrak{T}^\nu_\sigma$ the components of the energy of matter, $\mathfrak{t}^\nu_\sigma$ the components of the energy of the gravitational field.

Having regard to (20), there follows from the field equations (7) of gravitation, by multiplication by $g^{\mu\nu}_\sigma$ and summation with respect to μ and ν,

$$\frac{\partial \mathfrak{t}^\nu_\sigma}{\partial x_\nu} + \tfrac{1}{2}g^{\mu\nu}_\sigma\frac{\partial \mathfrak{M}}{\partial g^{\mu\nu}} = 0,$$

or, in view of (19) and (21),

$$\frac{\partial \mathfrak{T}_\sigma^\nu}{\partial x_\nu} + \tfrac{1}{2} g_\sigma^{\mu\nu} \mathfrak{T}_{\mu\nu} = 0 \quad . \quad . \quad . \quad (22)$$

where $\mathfrak{T}_{\mu\nu}$ denotes the quantities $g_{\nu\sigma}\mathfrak{T}_\mu^\sigma$. These are four equations which the energy-components of matter have to satisfy.

It is to be emphasized that the (generally covariant) laws of conservation (21) and (22) are deduced from the field equations (7) of gravitation, in combination with the postulate of general covariance (relativity) *alone*, without using the field equations (8) for material phenomena.

COSMOLOGICAL CONSIDERATIONS ON THE GENERAL THEORY OF RELATIVITY

BY

A. EINSTEIN

Translated from "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie," Sitzungsberichte der Preussischen Akad. d. Wissenschaften, 1917.

COSMOLOGICAL CONSIDERATIONS ON THE GENERAL THEORY OF RELATIVITY

By A. EINSTEIN

IT is well known that Poisson's equation

$$\nabla^2\phi = 4\pi K\rho \quad . \quad . \quad . \quad . \quad (1)$$

in combination with the equations of motion of a material point is not as yet a perfect substitute for Newton's theory of action at a distance. There is still to be taken into account the condition that at spatial infinity the potential ϕ tends toward a fixed limiting value. There is an analogous state of things in the theory of gravitation in general relativity. Here, too, we must supplement the differential equations by limiting conditions at spatial infinity, if we really have to regard the universe as being of infinite spatial extent.

In my treatment of the planetary problem I chose these limiting conditions in the form of the following assumption: it is possible to select a system of reference so that at spatial infinity all the gravitational potentials $g_{\mu\nu}$ become constant. But it is by no means evident *a priori* that we may lay down the same limiting conditions when we wish to take larger portions of the physical universe into consideration. In the following pages the reflexions will be given which, up to the present, I have made on this fundamentally important question.

§ 1. The Newtonian Theory

It is well known that Newton's limiting condition of the constant limit for ϕ at spatial infinity leads to the view that the density of matter becomes zero at infinity. For we imagine that there may be a place in universal space round about which the gravitational field of matter, viewed on a large scale, possesses spherical symmetry. It then follows from Poisson's equation that, in order that ϕ may tend to a limit at infinity, the mean density ρ must decrease toward zero more rapidly than $1/r^2$ as the distance r from the centre increases.* In this sense, therefore, the universe according to Newton is finite, although it may possess an infinitely great total mass.

From this it follows in the first place that the radiation emitted by the heavenly bodies will, in part, leave the Newtonian system of the universe, passing radially outwards, to become ineffective and lost in the infinite. May not entire heavenly bodies fare likewise? It is hardly possible to give a negative answer to this question. For it follows from the assumption of a finite limit for ϕ at spatial infinity that a heavenly body with finite kinetic energy is able to reach spatial infinity by overcoming the Newtonian forces of attraction. By statistical mechanics this case must occur from time to time, as long as the total energy of the stellar system—transferred to one single star—is great enough to send that star on its journey to infinity, whence it never can return.

We might try to avoid this peculiar difficulty by assuming a very high value for the limiting potential at infinity. That would be a possible way, if the value of the gravitational potential were not itself necessarily conditioned by the heavenly bodies. The truth is that we are compelled to regard the occurrence of any great differences of potential of the gravitational field as contradicting the facts. These differences must really be of so low an order of magnitude that the stellar velocities generated by them do not exceed the velocities actually observed.

If we apply Boltzmann's law of distribution for gas molecules to the stars, by comparing the stellar system with a gas in thermal equilibrium, we find that the Newtonian stellar system cannot exist at all. For there is a finite ratio of densities corresponding to the finite difference of potential between the centre and spatial infinity. A vanishing of the density at infinity thus implies a vanishing of the density at the centre.

* ρ is the mean density of matter, calculated for a region which is large as compared with the distance between neighbouring fixed stars, but small in comparison with the dimensions of the whole stellar system.

It seems hardly possible to surmount these difficulties on the basis of the Newtonian theory. We may ask ourselves the question whether they can be removed by a modification of the Newtonian theory. First of all we will indicate a method which does not in itself claim to be taken seriously; it merely serves as a foil for what is to follow. In place of Poisson's equation we write

$$\nabla^2\phi - \lambda\phi = 4\pi\kappa\rho \quad . \quad . \quad . \quad (2)$$

where λ denotes a universal constant. If ρ_0 be the uniform density of a distribution of mass, then

$$\phi = -\frac{4\pi\kappa}{\lambda}\rho_0 \quad . \quad . \quad . \quad . \quad (3)$$

is a solution of equation (2). This solution would correspond to the case in which the matter of the fixed stars was distributed uniformly through space, if the density ρ_0 is equal to the actual mean density of the matter in the universe. The solution then corresponds to an infinite extension of the central space, filled uniformly with matter. If, without making any change in the mean density, we imagine matter to be non-uniformly distributed locally, there will be, over and above the ϕ with the constant value of equation (3), an additional ϕ, which in the neighbourhood of denser masses will so much the more resemble the Newtonian field as $\lambda\phi$ is smaller in comparison with $4\pi\kappa\rho$.

A universe so constituted would have, with respect to its gravitational field, no centre. A decrease of density in spatial infinity would not have to be assumed, but both the mean potential and mean density would remain constant to infinity. The conflict with statistical mechanics which we found in the case of the Newtonian theory is not repeated. With a definite but extremely small density, matter is in equilibrium, without any internal material forces (pressures) being required to maintain equilibrium.

§ 2. The Boundary Conditions According to the General Theory of Relativity

In the present paragraph I shall conduct the reader over the road that I have myself travelled, rather a rough and winding road, because otherwise I cannot hope that he will take much interest in the result at the end of the journey. The conclusion I shall arrive at is that the field equations of gravitation which I have championed hitherto still need a slight modification, so that on the basis of the general theory of relativity those fundamental difficulties may be avoided which have been set forth in § 1 as confronting the Newtonian theory. This modification corresponds perfectly to the transition from Poisson's equation (1) to equation (2) of § 1. We finally infer that boundary conditions in spatial infinity fall away altogether, because the universal continuum in respect of its spatial dimensions is to be viewed as a self-contained continuum of finite spatial (three-dimensional) volume.

The opinion which I entertained until recently, as to the limiting conditions to be laid down in spatial infinity, took its stand on the following considerations. In a consistent theory of relativity there can be no inertia *relatively to "space,"* but only an inertia of masses *relatively to one another*. If, therefore, I have a mass at a sufficient distance from all other masses in the universe, its inertia must fall to zero. We will try to formulate this condition mathematically.

According to the general theory of relativity the negative momentum is given by the first three components, the energy by the last component of the covariant tensor multiplied by $\sqrt{-g}$

$$m\sqrt{-g} \quad g_{\mu\alpha}\frac{dx_\alpha}{ds} \quad . \quad . \quad . \quad . \quad (4)$$

where, as always, we set

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu \quad . \quad . \quad . \quad . \quad (5)$$

In the particularly perspicuous case of the possibility of choosing the system of co-ordinates so that the gravitational field at every point is spatially isotropic, we have more simply

$$ds^2 = -\mathrm{A}(dx_1^2 + dx_2^2 + dx_3^2) + \mathrm{B}dx_4^2.$$

If, moreover, at the same time

$$\sqrt{-g} = 1 = \sqrt{\mathrm{A}^3\mathrm{B}}$$

we obtain from (4), to a first approximation for small velocities;

$$m\frac{\mathrm{A}}{\sqrt{\mathrm{B}}}\frac{dx_1}{dx_4},\ m\frac{\mathrm{A}}{\sqrt{\mathrm{B}}}\frac{dx_2}{dx_4},\ m\frac{\mathrm{A}}{\sqrt{\mathrm{B}}}\frac{dx_3}{dx_4}$$

for the components of momentum, and for the energy (in the static case)

$$m\sqrt{B}.$$

From the expressions for the momentum, it follows that $m\frac{A}{\sqrt{B}}$ plays the part of the rest mass. As m is a constant peculiar to the point of mass, independently of its position, this expression, if we retain the condition $\sqrt{g -} = 1$ at spatial infinity, can vanish only when A diminishes to zero, while B increases to infinity. It seems, therefore, that such a degeneration of the co-efficients $g_{\mu\nu}$ is required by the postulate of relativity of all inertia. This requirement implies that the potential energy $m\sqrt{B}$ becomes infinitely great at infinity. Thus a point of mass can never leave the system; and a more detailed investigation shows that the same thing applies to light-rays. A system of the universe with such behaviour of the gravitational potentials at infinity would not therefore run the risk of wasting away which was mooted just now in connexion with the Newtonian theory.

I wish to point out that the simplifying assumptions as to the gravitational potentials on which this reasoning is based, have been introduced merely for the sake of lucidity. It is possible to find general formulations for the behaviour of the $g_{\mu\nu}$ at infinity which express the essentials of the question without further restrictive assumptions.

At this stage, with the kind assistance of the mathematician J. Grommer, I investigated centrally symmetrical, static gravitational fields, degenerating at infinity in the way mentioned. The gravitational potentials $g_{\mu\nu}$ were applied, and from them the energy-tensor $T_{\mu\nu}$ of matter was calculated on the basis of the field equations of gravitation. But here it proved that for the system of the fixed stars no boundary conditions of the kind can come into question at all, as was also rightly emphasized by the astronomer de Sitter recently.

For the contravariant energy-tensor $T^{\mu\nu}$ of ponderable matter is given by

$$T^{\mu\nu} = \rho\frac{dx_\mu}{ds}\frac{dx_\nu}{ds},$$

where ρ is the density of matter in natural measure. With an appropriate choice of the system of co-ordinates the stellar velocities are very small in comparison with that of light. We may, therefore, substitute $\sqrt{g_{44}}\, dx_4$ for ds. This shows us that all components of $T^{\mu\nu}$ must be very small in comparison with the last component T^{44}. But it was quite impossible to reconcile this condition with the chosen boundary conditions. In the retrospect this result does not appear astonishing. The fact of the small velocities of the stars allows the conclusion that wherever there are fixed stars, the gravitational potential (in our case $\sqrt{B}$) can never be much greater than here on earth. This follows from statistical reasoning, exactly as in the case of the Newtonian theory. At any rate, our calculations have convinced me that such conditions of degeneration for the $g_{\mu\nu}$ in spatial infinity may not be postulated.

After the failure of this attempt, two possibilities next present themselves.

(*a*) We may require, as in the problem of the planets, that, with a suitable choice of the system of reference, the $g_{\mu\nu}$ in spatial infinity approximate to the values

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

(*b*) We may refrain entirely from laying down boundary conditions for spatial infinity claiming general validity; but at the spatial limit of the domain under consideration we have to give the $g_{\mu\nu}$ separately in each individual case, as hitherto we were accustomed to give the initial conditions for time separately.

The possibility (*b*) holds out no hope of solving the problem, but amounts to giving it up. This is an incontestable position, which is taken up at the present time by de Sitter.* But I must confess that such a complete resignation in this fundamental question is for me a difficult thing. I should not make up my mind to it until every effort to make headway toward a satisfactory view had proved to be vain.

Possibility (*a*) is unsatisfactory in more respects than one.

* de Sitter, Akad. van Wetensch. te Amsterdam, 8 Nov., 1916.

In the first place those boundary conditions pre-suppose a definite choice of the system of reference, which is contrary to the spirit of the relativity principle. Secondly, if we adopt this view, we fail to comply with the requirement of the relativity of inertia. For the inertia of a material point of mass m (in natural measure) depends upon the $g_{\mu\nu}$; but these differ but little from their postulated values, as given above, for spatial infinity. Thus inertia would indeed be *influenced*, but would not be *conditioned* by matter (present in finite space). If only one single point of mass were present, according to this view, it would possess inertia, and in fact an inertia almost as great as when it is surrounded by the other masses of the actual universe. Finally, those statistical objections must be raised against this view which were mentioned in respect of the Newtonian theory.

From what has now been said it will be seen that I have not succeeded in formulating boundary conditions for spatial infinity. Nevertheless, there is still a possible way out, without resigning as suggested under (*b*). For if it were possible to regard the universe as a continuum which is *finite* (*closed*) *with respect to its spatial dimensions*, we should have no need at all of any such boundary conditions. We shall proceed to show that both the general postulate of relativity and the fact of the small stellar velocities are compatible with the hypothesis of a spatially finite universe; though certainly, in order to carry through this idea, we need a generalizing modification of the field equations of gravitation.

§ 3. The Spatially Finite Universe with a Uniform Distribution of Matter

According to the general theory of relativity the metrical character (curvature) of the four-dimensional space-time continuum is defined at every point by the matter at that point and the state of that matter. Therefore, on account of the lack of uniformity in the distribution of matter, the metrical structure of this continuum must necessarily be extremely complicated. But if we are concerned with the structure only on a large scale, we may represent matter to ourselves as being uniformly distributed over enormous spaces, so that its density of distribution is a variable function which varies extremely slowly. Thus our procedure will somewhat resemble that of the geodesists who, by means of an ellipsoid, approximate to the shape of the earth's surface, which on a small scale is extremely complicated.

The most important fact that we draw from experience as to the distribution of matter is that the relative velocities of the stars are very small as compared with the velocity of light. So I think that for the present we may base our reasoning upon the following approximative assumption. There is a system of reference relatively to which matter may be looked upon as being permanently at rest. With respect to this system, therefore, the contravariant energy-tensor $T^{\mu\nu}$ of matter is, by reason of (5), of the simple form

$$\left.\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho \end{array}\right\} \quad . \quad . \quad . \quad (6)$$

The scalar ρ of the (mean) density of distribution may be *a priori* a function of the space co-ordinates. But if we assume the universe to be spatially finite, we are prompted to the hypothesis that ρ is to be independent of locality. On this hypothesis we base the following considerations.

As concerns the gravitational field, it follows from the equation of motion of the material point

$$\frac{d^2x_\nu}{ds^2} + \{\alpha\beta, \nu\}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 0$$

that a material point in a static gravitational field can remain at rest only when g_{44} is independent of locality. Since, further, we presuppose independence of the time co-ordinate x_4 for all magnitudes, we may demand for the required solution that, for all x_ν,

$$g_{44} = 1 \quad . \quad . \quad . \quad . \quad (7)$$

Further, as always with static problems, we shall have to set

$$g_{14} = g_{24} = g_{34} = 0 \quad . \quad . \quad . \quad (8)$$

It remains now to determine those components of the gravitational potential which define the purely spatial-geometrical relations of our continuum ($g_{11}, g_{12}, \ldots g_{33}$). From our assumption as to the uniformity of distribution of the masses generating the field, it follows that the curvature of the required space must be constant. With this distribution of mass, therefore, the required finite continuum of the x_1, x_2, x_3, with constant x_4, will be a spherical space.

We arrive at such a space, for example, in the following way. We start from a Euclidean space of four dimensions, ξ_1, ξ_2, ξ_3 ξ_4, with a linear element $d\sigma$; let, therefore,

$$d\sigma^2 = d\xi_1^2 + d\xi_2^2 + d\xi_3^2 + d\xi_4^2 \quad . \quad . \quad . \quad (9)$$

In this space we consider the hyper-surface

$$R^2 = \xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2, \quad . \quad . \quad . \quad (10)$$

where R denotes a constant. The points of this hyper-surface form a three-dimensional continuum, a spherical space of radius of curvature R.

The four-dimensional Euclidean space with which we started serves only for a convenient definition of our hyper-surface. Only those points of the hyper-surface are of interest to us which have metrical properties in agreement with those of physical space with a uniform distribution of matter. For the description of this three-dimensional continuum we may employ the co-ordinates ξ_1, ξ_2, ξ_3 (the projection upon the hyper-plane $\xi_4 = 0$) since, by reason of (10), ξ_4 can be expressed in terms of ξ_1, ξ_2, ξ_3. Eliminating ξ_4 from (9), we obtain for the linear element of the spherical space the expression

$$\left.\begin{aligned} d\sigma^2 &= \gamma_{\mu\nu} d\xi_\mu d\xi_\nu \\ \gamma_{\mu\nu} &= \delta_{\mu\nu} + \frac{\xi_\mu \xi_\nu}{R^2 - \rho^2} \end{aligned}\right\} \quad . \quad . \quad . \quad (11)$$

where $\delta_{\mu\nu} = 1$, if $\mu = \nu$; $\delta_{\mu\nu} = 0$, if $\mu \neq \nu$, and $\rho^2 = \xi_1^2 + \xi_2^2 + \xi_3^2$. The co-ordinates chosen are convenient when it is a question of examining the environment of one of the two points $\xi_1 = \xi_2 = \xi_3 = 0$.

Now the linear element of the required four-dimensional space-time universe is also given us. For the potential $g_{\mu\nu}$, both indices of which differ from 4, we have to set

$$g_{\mu\nu} = -\left(\delta_{\mu\nu} + \frac{x_\mu x_\nu}{R^2 - (x_1^2 + x_2^2 + x_3^2)}\right) \quad . \quad (12)$$

which equation, in combination with (7) and (8), perfectly defines the behaviour of measuring-rods, clocks, and light-rays.

§ 4. On an Additional Term for the Field Equations of Gravitation

My proposed field equations of gravitation for any chosen system of co-ordinates run as follows :—

$$\left.\begin{aligned} G_{\mu\nu} &= -\kappa(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T), \\ G_{\mu\nu} &= -\frac{\partial}{\partial x_\alpha}\{\mu\nu, \alpha\} + \{\mu\alpha, \beta\}\{\nu\beta, \alpha\} \\ &\qquad + \frac{\partial^2 \log\sqrt{-g}}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \alpha\}\frac{\partial \log\sqrt{-g}}{\partial x_\alpha} \end{aligned}\right\} \quad (13)$$

The system of equations (13) is by no means satisfied when we insert for the $g_{\mu\nu}$ the values given in (7), (8), and (12), and for the (contravariant) energy-tensor of matter the values indicated in (6). It will be shown in the next paragraph how this calculation may conveniently be made. So that, if it were certain that the field equations (13) which I have hitherto employed were the only ones compatible with the postulate of general relativity, we should probably have to conclude that the theory of relativity does not admit the hypothesis of a spatially finite universe.

However, the system of equations (14) allows a readily suggested extension which is compatible with the relativity postulate, and is perfectly analogous to the extension of Poisson's equation given by equation (2). For on the left-hand side of field equation (13) we may add the fundamental tensor $g_{\mu\nu}$, multiplied by a universal constant, $-\lambda$, at present unknown, without destroying the general covariance. In place of field equation (13) we write

$$G_{\mu\nu} - \lambda g_{\mu\nu} = -\kappa(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T) \quad . \quad . \quad (13a)$$

This field equation, with λ sufficiently small, is in any case also compatible with the facts of experience derived from the solar system. It also satisfies laws of conservation of momentum and energy, because we arrive at (13a) in place of (13) by introducing into Hamilton's principle, instead of the scalar of Riemann's tensor, this scalar increased by a universal constant; and Hamilton's principle, of course, guarantees the validity of laws of conservation. It will be shown in § 5 that field equation (13a) is compatible with our conjectures on field and matter.

§ 5. Calculation and Result

Since all points of our continuum are on an equal footing, it is sufficient to carry through the calculation for *one* point, e.g. for one of the two points with the co-ordinates

$$x_1 = x_2 = x_3 = x_4 = 0.$$

Then for the $g_{\mu\nu}$ in (13a) we have to insert the values

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

wherever they appear differentiated only once or not at all. We thus obtain in the first place

$$G_{\mu\nu} = \frac{\partial}{\partial x_1}[\mu\nu, 1] + \frac{\partial}{\partial x_2}[\mu\nu, 2] + \frac{\partial}{\partial x_3}[\mu\nu, 3] + \frac{\partial^2 \log \sqrt{-g}}{\partial x_\mu \partial x_\nu}.$$

From this we readily discover, taking (7), (8), and (13) into account, that all equations (13*a*) are satisfied if the two relations

$$-\frac{2}{R^2} + \lambda = -\frac{\kappa\rho}{2}, \quad -\lambda = -\frac{\kappa\rho}{2},$$

or

$$\lambda = \frac{\kappa\rho}{2} = \frac{1}{R^2} \quad . \quad . \quad . \quad . \quad (14)$$

are fulfilled.

Thus the newly introduced universal constant λ defines both the mean density of distribution ρ which can remain in equilibrium and also the radius R and the volume $2\pi^2 R^3$ of spherical space. The total mass M of the universe, according to our view, is finite, and is in fact

$$M = \rho \, . \, 2\pi^2 R^3 = 4\pi^2 \frac{R}{\kappa} = \pi^2 \sqrt{\frac{32}{\kappa^3 \rho}} \quad . \quad . \quad (15)$$

Thus the theoretical view of the actual universe, if it is in correspondence with our reasoning, is the following. The curvature of space is variable in time and place, according to the distribution of matter, but we may roughly approximate to it by means of a spherical space. At any rate, this view is logically consistent, and from the standpoint of the general theory of relativity lies nearest at hand; whether, from the standpoint of present astronomical knowledge, it is tenable, will not here be discussed. In order to arrive at this consistent view, we admittedly had to introduce an extension of the field equations of gravitation which is not justified by our actual knowledge of gravitation. It is to be emphasized, however, that a positive curvature of space is given by our results, even if the supplementary term is not introduced. That term is necessary only for the purpose of making possible a quasi-static distribution of matter, as required by the fact of the small velocities of the stars.

DO GRAVITATIONAL FIELDS PLAY AN ESSENTIAL PART IN THE STRUCTURE OF THE ELEMENTARY PARTICLES OF MATTER?

BY

A. EINSTEIN

Translated from " Spielen Gravitationsfelder im Aufber der materiellen Elementarteilchen eine wesentliche Rolle?" Sitzungsberichte der Preussischen Akad. d. Wissenschaften, 1919.

DO GRAVITATIONAL FIELDS PLAY AN ESSENTIAL PART IN THE STRUCTURE OF THE ELEMENTARY PARTICLES OF MATTER?

By A. EINSTEIN

NEITHER the Newtonian nor the relativistic theory of gravitation has so far led to any advance in the theory of the constitution of matter. In view of this fact it will be shown in the following pages that there are reasons for thinking that the elementary formations which go to make up the atom are held together by gravitational forces.

§ 1. *Defects of the Present View*

Great pains have been taken to elaborate a theory which will account for the equilibrium of the electricity constituting the electron. G. Mie, in particular, has devoted deep researches to this question. His theory, which has found considerable support among theoretical physicists, is based mainly on the introduction into the energy-tensor of supplementary terms depending on the components of the electro-dynamic potential, in addition to the energy terms of the Maxwell-Lorentz theory. These new terms, which in outside space are unimportant, are nevertheless effective in the interior of the electrons in maintaining equilibrium against the electric forces of repulsion. In spite of the beauty of the formal structure of this theory, as erected by Mie, Hilbert, and Weyl, its physical results have hitherto been unsatisfactory. On the one hand the multiplicity of possibilities is discouraging, and on the other hand those additional terms have not as yet allowed themselves to be framed in such a simple form that the solution could be satisfactory.

So far the general theory of relativity has made no change in this state of the question. If we for the moment disregard the additional cosmological term, the field equations take the form

$$G_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}G = -\kappa T_{\mu\nu} \qquad . \quad . \quad . \quad (1)$$

where $G_{\mu\nu}$ denotes the contracted Riemann tensor of curvature, G the scalar of curvature formed by repeated contraction, and $T_{\mu\nu}$ the energy-tensor of "matter." The assumption that the $T_{\mu\nu}$ do *not* depend on the derivatives of the $g_{\mu\nu}$ is in keeping with the historical development of these equations. For these quantities are, of course, the energy-components in the sense of the special theory of relativity, in which variable $g_{\mu\nu}$ do not occur. The second term on the left-hand side of the equation is so chosen that the divergence of the left-hand side of (1) vanishes identically, so that taking the divergence of (1), we obtain the equation

$$\frac{\partial \mathfrak{T}^{\sigma}_{\mu}}{\partial x_{\sigma}} + \tfrac{1}{2}g^{\sigma\tau}_{\mu}\mathfrak{T}_{\sigma\tau} = 0 \qquad . \quad . \quad . \quad (2)$$

which in the limiting case of the special theory of relativity gives the complete equations of conservation

$$\frac{\partial T_{\mu\nu}}{\partial x_{\nu}} = 0.$$

Therein lies the physical foundation for the second term of the left-hand side of (1). It is by no means settled *a priori* that a limiting transition of this kind has any possible meaning. For if gravitational fields do play an essential part in the structure of the particles of matter, the transition to the limiting case of constant $g_{\mu\nu}$ would, for them, lose its justification, for indeed, with constant $g_{\mu\nu}$ there could not be any particles of matter. So if we wish to contemplate the possibility that gravitation may take part in the structure of the fields which constitute the corpuscles, we cannot regard equation (1) as confirmed.

Placing in (1) the Maxwell-Lorentz energy-components of the electromagnetic field $\phi_{\mu\nu}$,

$$T_{\mu\nu} = \tfrac{1}{4}g_{\mu\nu}\phi_{\sigma\tau}\phi^{\sigma\tau} - \phi_{\mu\sigma}\phi_{\nu\tau}g^{\sigma\tau}, \qquad . \quad . \quad (3)$$

we obtain for (2), by taking the divergence, and after some reduction,*

$$\phi_{\mu\sigma}\mathfrak{J}^{\sigma} = 0 \quad . \quad . \quad . \quad . \quad (4)$$

where, for brevity, we have set

$$\frac{\partial}{\partial x_{\tau}}(\sqrt{-g}\ \phi_{\mu\nu}g^{\mu\sigma}g^{\nu\tau}) = \frac{\partial \mathfrak{f}^{\sigma\tau}}{\partial x_{\tau}} = \mathfrak{J}^{\sigma} \quad . \quad . \quad (5)$$

In the calculation we have employed the second of Maxwell's systems of equations

$$\frac{\partial \phi_{\mu\nu}}{\partial x_{\rho}} + \frac{\partial \phi_{\nu\rho}}{\partial x_{\mu}} + \frac{\partial \phi_{\rho\mu}}{\partial x_{\nu}} = 0 \quad . \quad . \quad . \quad (6)$$

We see from (4) that the current-density $\mathfrak{J}^{\sigma}$ must everywhere vanish. Therefore, by equation (1), we cannot arrive at a theory of the electron by restricting ourselves to the electromagnetic components of the Maxwell-Lorentz theory, as has long been known. Thus if we hold to (1) we are driven on to the path of Mie's theory.†

Not only the problem of matter, but the cosmological problem as well, leads to doubt as to equation (1). As I have shown in the previous paper, the general theory of relativity requires that the universe be spatially finite. But this view of the universe necessitated an extension of equations (1), with the introduction of a new universal constant λ, standing in a fixed relation to the total mass of the universe (or, respectively, to the equilibrium density of matter). This is gravely detrimental to the formal beauty of the theory.

§ 2. The Field Equations Freed of Scalars

The difficulties set forth above are removed by setting in place of field equations (1) the field equations

$$G_{\mu\nu} - \tfrac{1}{4}g_{\mu\nu}G = -\kappa T_{\mu\nu} \quad . \quad . \quad . \quad (1a)$$

where $T_{\mu\nu}$ denotes the energy-tensor of the electromagnetic field given by (3).

The formal justification for the factor $-\frac{1}{4}$ in the second

* Cf. e.g. A. Einstein, Sitzungsber. d. Preuss. Akad. d. Wiss., 1916, pp. 187, 188.

† Cf. D. Hilbert, Göttinger Nachr., 20 Nov., 1915.

term of this equation lies in its causing the scalar of the left-hand side,

$$g^{\mu\nu}(G_{\mu\nu} - \tfrac{1}{4}g_{\mu\nu}G),$$

to vanish identically, as the scalar $g^{\nu\mu}T_{\mu\nu}$ of the right-hand side does by reason of (3). If we had reasoned on the basis of equations (1) instead of (1a), we should, on the contrary, have obtained the condition $G = 0$, which would have to hold good everywhere for the $g_{\mu\nu}$, independently of the electric field. It is clear that the system of equations [(1a), (3)] is a consequence of the system [(1), (3)], but not conversely.

We might at first sight feel doubtful whether (1a) together with (6) sufficiently define the entire field. In a generally relativistic theory we need $n - 4$ differential equations, independent of one another, for the definition of n independent variables, since in the solution, on account of the liberty of choice of the co-ordinates, four quite arbitrary functions of all co-ordinates must naturally occur. Thus to define the sixteen independent quantities $g_{\mu\nu}$ and $\phi_{\mu\nu}$ we require twelve equations, all independent of one another. But as it happens, nine of the equations (1a), and three of the equations (6) are independent of one another.

Forming the divergence of (1a), and taking into account that the divergence of $G_{\mu\nu} - \frac{1}{2}g_{\mu\nu}G$ vanishes, we obtain

$$\phi_{\sigma\alpha}J^{\alpha} + \frac{1}{4\kappa}\frac{\partial G}{\partial x_{\sigma}} = 0 \quad . \quad . \quad . \quad (4a)$$

From this we recognize first of all that the scalar of curvature G in the four-dimensional domains in which the density of electricity vanishes, is constant. If we assume that all these parts of space are connected, and therefore that the density of electricity differs from zero only in separate "world-threads," then the scalar of curvature, everywhere outside these world-threads, possesses a constant value G_0. But equation (4a) also allows an important conclusion as to the behaviour of G within the domains having a density of electricity other than zero. If, as is customary, we regard electricity as a moving density of charge, by setting

$$J^{\sigma} = \frac{\mathfrak{J}^{\sigma}}{\sqrt{-g}} = \rho\frac{dx_{\sigma}}{ds}, \quad . \quad . \quad . \quad (7)$$

we obtain from (4a) by inner multiplication by J^σ, on account of the antisymmetry of $\phi_{\mu\nu}$, the relation

$$\frac{\partial G}{\partial x_\sigma}\frac{dx_\sigma}{ds} = 0 \quad . \quad . \quad . \quad . \quad (8)$$

Thus the scalar of curvature is constant on every world-line of the motion of electricity. Equation (4a) can be interpreted in a graphic manner by the statement: The scalar of curvature plays the part of a negative pressure which, outside of the electric corpuscles, has a constant value G_0. In the interior of every corpuscle there subsists a negative pressure (positive $G - G_0$) the fall of which maintains the electrodynamic force in equilibrium. The minimum of pressure, or, respectively, the maximum of the scalar of curvature, does not change with time in the interior of the corpuscle.

We now write the field equations (1a) in the form

$$(G_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}G) + \tfrac{1}{4}g_{\mu\nu}G_0 = -\kappa\left(T_{\mu\nu} + \frac{1}{4\kappa}g_{\mu\nu}(G - G_0)\right) \quad (9)$$

On the other hand, we transform the equations supplied with the cosmological term as already given

$$G_{\mu\nu} - \lambda g_{\mu\nu} = -\kappa(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T).$$

Subtracting the scalar equation multiplied by $\frac{1}{2}$, we next obtain

$$(G_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}G) + g_{\mu\nu}\lambda = -\kappa T_{\mu\nu}.$$

Now in regions where only electrical and gravitational fields are present, the right-hand side of this equation vanishes. For such regions we obtain, by forming the scalar,

$$-G + 4\lambda = 0.$$

In such regions, therefore, the scalar of curvature is constant, so that λ may be replaced by $\frac{1}{4}G_0$. Thus we may write the earlier field equation (1) in the form

$$G_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}G + \tfrac{1}{4}g_{\mu\nu}G_0 = -\kappa T_{\mu\nu} \quad . \quad . \quad (10)$$

Comparing (9) with (10), we see that there is no difference between the new field equations and the earlier ones, except that instead of $T_{\mu\nu}$ as tensor of "gravitating mass" there now occurs $T_{\mu\nu} + \frac{1}{4\kappa} g_{\mu\nu}(G - G_0)$ which is independent of the scalar of curvature. But the new formulation has this great advantage, that the quantity λ appears in the fundamental equations as a constant of integration, and no longer as a universal constant peculiar to the fundamental law.

§ 3. On the Cosmological Question

The last result already permits the surmise that with our new formulation the universe may be regarded as spatially finite, without any necessity for an additional hypothesis. As in the preceding paper I shall again show that with a uniform distribution of matter, a spherical world is compatible with the equations.

In the first place we set

$$ds^2 = -\gamma_{ik}dx_i dx_k + dx_4^2 \; (i, k = 1, 2, 3) \tag{11}$$

Then if P_{ik} and P are, respectively, the curvature tensor of the second rank and the curvature scalar in three-dimensional space, we have

$$\begin{aligned} G_{ik} &= P_{ik} \; (i, k = 1, 2, 3) \\ G_{i4} &= G_{4i} = G_{44} = 0 \\ G &= -P \\ -g &= \gamma. \end{aligned}$$

It therefore follows for our case that

$$\begin{aligned} G_{ik} - \tfrac{1}{2}g_{ik}G &= P_{ik} - \tfrac{1}{2}\gamma_{ik}P \; (i, k = 1, 2, 3) \\ G_{44} - \tfrac{1}{2}g_{44}G &= \tfrac{1}{2}P. \end{aligned}$$

We pursue our reflexions, from this point on, in two ways. Firstly, with the support of equation (1a). Here $T_{\mu\nu}$ denotes the energy-tensor of the electro-magnetic field, arising from the electrical particles constituting matter. For this field we have everywhere

$$\mathfrak{T}_1^1 + \mathfrak{T}_2^2 + \mathfrak{T}_3^3 + \mathfrak{T}_4^4 = 0.$$

The individual $\mathfrak{T}_\mu^\nu$ are quantities which vary rapidly with position; but for our purpose we no doubt may replace them by their mean values. We therefore have to choose

$$\left.\begin{aligned} \mathfrak{T}_1^1 &= \mathfrak{T}_2^2 = \mathfrak{T}_3^3 = -\tfrac{1}{3}\mathfrak{T}_4^4 = \text{const.} \\ \mathfrak{T}_\mu^\nu &= 0 \;\; (\text{for } \mu \neq \nu), \end{aligned}\right\} \tag{12}$$

and therefore

$$T_{ik} = \frac{1}{3}\frac{\mathfrak{T}_4^4}{\sqrt{\gamma}}\gamma_{ik}, \quad T_{44} = \frac{\mathfrak{T}_4^4}{\sqrt{\gamma}}.$$

In consideration of what has been shown hitherto, we obtain in place of (1a)

$$P_{ik} - \frac{1}{4}\gamma_{ik}P = -\frac{1}{3}\gamma_{ik}\frac{\kappa\mathfrak{T}_4^4}{\sqrt{\gamma}} \quad . \quad . \quad . \quad (13)$$

$$\frac{1}{4}P = -\frac{\kappa\mathfrak{T}_4^4}{\sqrt{\gamma}} \quad . \quad . \quad . \quad (14)$$

The scalar of equation (13) agrees with (14). It is on this account that our fundamental equations permit the idea of a spherical universe. For from (13) and (14) follows

$$P_{ik} + \frac{4}{3}\frac{\kappa\mathfrak{T}_4^4}{\sqrt{\gamma}}\gamma_{ik} = 0 \quad . \quad . \quad . \quad (15)$$

and it is known * that this system is satisfied by a (three-dimensional) spherical universe.

But we may also base our reflexions on the equations (9). On the right-hand side of (9) stand those terms which, from the phenomenological point of view, are to be replaced by the energy-tensor of matter; that is, they are to be replaced by

$$\begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho \end{matrix}$$

where ρ denotes the mean density of matter assumed to be at rest. We thus obtain the equations

$$P_{ik} - \frac{1}{2}\gamma_{ik}P - \frac{1}{4}\gamma_{ik}G_0 = 0 \quad . \quad . \quad (16)$$

$$\frac{1}{2}P + \frac{1}{4}G_0 = -\kappa\rho \quad . \quad . \quad . \quad . \quad (17)$$

From the scalar of equation (16) and from (17) we obtain

$$G_0 = -\frac{2}{3}P = 2\kappa\rho, \quad . \quad . \quad . \quad (18)$$

and consequently from (16)

$$P_{ik} - \kappa\rho\gamma_{ik} = 0 \quad . \quad . \quad . \quad (19)$$

* Cf. H. Weyl, "Raum, Zeit, Materie," § 33.

GRAVITATION AND ELECTRICITY*

By H. WEYL

ACCORDING to Riemann,† geometry is based upon the following two facts :—

1. *Space is a Three-dimensional Continuum.*—The manifold of its points may therefore be consistently represented by the values of three co-ordinates x_1, x_2, x_3.

2. (*Pythagorean Theorem*).—The square of the distance ds between two infinitely proximate points

$P = (x_1, x_2, x_3)$ and $P' = (x_1 + dx_1, x_2 + dx_2, x_3 + dx_3)$ (1)

(any co-ordinates being employed) is a quadratic form of the relative co-ordinates dx_μ :—

$$ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx_\mu dx_\nu, \quad (g_{\mu\nu} = g_{\nu\mu}) \quad . \quad . \quad (2)$$

The second of these facts may be briefly stated by saying that space is a *metrical* continuum. In complete accord with the spirit of the physics of immediate action we assume the Pythagorean theorem to be strictly valid only in the limit when the distances are infinitely small.

The special theory of relativity led to the discovery that *time* is associated as a fourth co-ordinate (x_4) on an equal footing with the three co-ordinates of space, and that the scene of material events, *the world*, is therefore *a four-dimensional, metrical continuum.* And so the quadratic form (2), which defines the metrical properties of the world, is not necessarily positive as in the case of the geometry of three-dimensional space, but has the index of inertia 3.‡ Riemann

* The footnotes in square brackets are later additions by the author.

† Math. Werke (2nd ed., Leipzig, 1892), No. XII, p. 282.

‡ That is to say that if the co-ordinates are chosen so that at one particular point of the continuum $ds^2 = \pm dx_1^2 \pm dx_2^2 \pm dx_3^2 \pm dx_4^2$, then in every case three of the signs will be + and one − (TRANS.).

himself did not fail to point out that this quadratic form was to be regarded as a physical reality, since it reveals itself, e.g. in centrifugal forces, as the origin of real effects upon matter, and that matter therefore presumably reacts upon it. Until then all geometricians and philosophers had looked upon the metrical properties of space as pertaining to space itself, independently of the matter which it contained. It is upon this idea, which it was quite impossible for Riemann in his day to carry through, that Einstein in our own time, independently of Riemann, has raised the imposing edifice of his general theory of relativity. According to Einstein the phenomena of *gravitation* must also be placed to the account of geometry, and the laws by which matter affects measurements are no other than the laws of gravitation: the $g_{\mu\nu}$ in (2) form the components of the gravitational potential. While the gravitational potential thus consists of an invariant *quadratic* differential form, *electromagnetic phenomena* are governed by a four-potential of which the components ϕ_μ together compose an invariant *linear* differential form $\Sigma\phi_\mu dx_\mu$. But so far the two classes of phenomena, gravitation and electricity, stand side by side, the one separate from the other.

The later work of Levi-Civita,* Hessenberg,† and the author‡ shows quite plainly that the fundamental conception on which the development of Riemann's geometry must be based if it is to be in agreement with nature, is that of the infinitesimal parallel displacement of a vector. If P and P* are any two points connected by a curve, a given vector at P can be moved parallel to itself along this curve from P to P*. But, generally speaking, this conveyance of a vector from P to P* is not integrable, that is to say, the vector at P* at which we arrive depends upon the path along which the displacement travels. It is only in Euclidean "gravitationless" geometry that integrability obtains. The Riemannian geometry referred to above still contains a residual element of finite geometry—without any substantial reason, as far as I can see.

* "Nozione di parallelismo . . .", Rend. del Circ. Matem. di Palermo, Vol. 42 (1917).

† "Vektorielle Begründung der Differentialgeometrie," Math. Ann., Vol. 78 (1917).

‡ "Space, Time, and Matter" (1st ed., Berlin, 1918), § 14.

It seems to be due to the accidental origin of this geometry in the theory of surfaces. The quadratic form (2) enables us to compare, with respect to their length, not only two vectors at the same point, but also the vectors at any two points. *But a truly infinitesimal geometry must recognize only the principle of the transference of a length from one point to another point infinitely near to the first.* This forbids us to assume that the problem of the transference of length from one point to another at a finite distance is integrable, more particularly as the problem of the transference of direction has proved to be non-integrable. Such an assumption being recognized as false, a geometry comes into being, which, when applied to the world, explains in a surprising manner *not only the phenomena of gravitation, but also those of the electromagnetic field.* According to the theory which now takes shape, both classes of phenomena spring from the same source, and in fact *we cannot in general make any arbitrary separation of electricity from gravitation.* In this theory *all physical quantities have a meaning in world geometry.* In particular the quantities denoting physical effects appear at once as pure numbers. The theory leads to a world-law which in its essentials is defined without ambiguity. It even permits us in a certain sense to comprehend why the world has four dimensions. I shall now first of all give a sketch of the structure of the amended geometry of Riemann without any thought of its physical interpretation. Its application to physics will then follow of its own accord.

In a given system of co-ordinates the relative co-ordinates dx_μ of a point P′ infinitely near to P—see (1)—are the components of the infinitesimal displacement PP′. The transition from one system of co-ordinates to another is expressed by definite formulæ of transformation,

$$x_\mu = x_\mu(x_1^*, x_2^* \ldots x_n^*) \quad \mu = 1, 2, \ldots n,$$

which determine the connexion between the co-ordinates of the same point in the two systems. Then between the components dx_μ and the components dx_μ^* of the same infinitesimal displacement of the point P we have the linear formulæ of transformation

$$dx_\mu = \sum_\nu a_{\mu\nu} dx_\nu^* \quad . \quad . \quad . \quad . \quad (3)$$

in which $a_{\mu\nu}$ are the values of the derivatives $\frac{\partial x_\mu}{\partial x^*_\nu}$ at the point P. A contravariant vector $\mathbf{x}$ at the point P referred to either system of co-ordinates has n known numbers ξ^μ for its components, which in the transition to another system are transformed in exactly the same way (3) as the components of an infinitesimal displacement. I denote the totality of vectors at the point P as the vector-space at P. It is, firstly, linear or affine, i.e. by multiplication of a vector at P by a number, and by addition of two such vectors, there always arises a vector at P; and, secondly, it is metrical, i.e. by the symmetrical bilinear form belonging to (2) a scalar product

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x} = \sum_{\mu\nu} g_{\mu\nu}\xi^\mu\eta^\nu$$

is invariantly assigned to each pair of vectors $\mathbf{x}$ and $\mathbf{y}$ with components ξ^μ, η^μ. We take it, however, that this form is determined only as far as to a positive factor of proportionality, which remains arbitrary. If the manifold of points of space is represented by co-ordinates x_μ, the $g_{\mu\nu}$ are determined by the metrical properties at the point P only to the extent of their proportionality. In the physical sense, too, it is only the ratios of the $g_{\mu\nu}$ that has an immediate tangible meaning. For the equation

$$\sum_{\mu\nu} g_{\mu\nu}dx_\mu dx_\nu = 0$$

is satisfied, when P is a given origin, by those infinitely proximate world-points which are reached by a light signal emitted at P. For the purpose of analytical presentation we have firstly to choose a definite system of co-ordinates, and secondly at each point P to determine the arbitrary factor of proportionality with which the $g_{\mu\nu}$ are endowed. Accordingly the formulæ which emerge must possess a double property of invariance: they must be invariant with respect to any continuous transformations of co-ordinates, and they must remain unaltered if $\lambda g_{\mu\nu}$, where λ is an arbitrary continuous function of position, is substituted for the $g_{\mu\nu}$. The supervention of this second property of invariance is characteristic of our theory.

If P, P* are any two points, and if to each vector $\mathbf{x}$ at P a vector $\mathbf{x}^*$ at P* is assigned in such a way that in general $a\mathbf{x}$

becomes $a\mathbf{x}^*$, and $\mathbf{x} + \mathbf{y}$ becomes $\mathbf{x}^* + \mathbf{y}^*$ (a being any assigned number), and the vector zero at P is the only one to which the vector zero at P* corresponds, we then have made an affine or linear replica of the vector-space at P on the vector-space at P*. This replica has a particularly close resemblance when the scalar product of the vectors $\mathbf{x}^*$, $\mathbf{y}^*$ at P* is proportional to that of $\mathbf{x}$ and $\mathbf{y}$ at P for all pairs of vectors $\mathbf{x}$, $\mathbf{y}$ (In our view it is only this idea of a similar replica that has an objective sense, the previous theory permitted the more definite conception of a congruent replica) The meaning of the parallel displacement of a vector at the point P to a neighbouring point P is settled by the two axiomatic postulates

1 By the parallel displacement of the vectors at the point P to the neighbouring point P a similar image of the vector-space at P is made upon the vector-space at P′

2 If P_1, P_2 are two points in the neighbourhood of P, and the infinitesimal vector PP_2 at P is transformed into P_1P_{12} by a parallel displacement to the point P_1, while PP_1 at P is transformed into P_2P_{21} by parallel displacement to P_2, then P_{12}, P_{21} coincide, i e infinitesimal parallel displacements are commutative

That part of postulate 1 which says that the parallel displacement is an affine transposition of the vector-space from P to P′, is expressed analytically as follows the vector ξ^μ at $P = (x_1, x_2, \quad x_n)$ is by displacement transformed into a vector $\xi^\mu + d\xi^\mu$ at $P = (x_1 + dx_1, x_2 + dx_2, \quad x_n + dx_n)$ the components of which are in a linear relation to ξ^μ,—

$$d\xi^\mu = -\sum_\nu d\gamma^\mu_\nu \xi^\nu \qquad . \qquad . \quad (4)$$

The second postulate teaches that the $d\gamma^\mu_\nu$ are linear differential forms

$$d\gamma^\mu_\nu = \sum_\rho \Gamma^\mu_{\nu\rho} dx_\rho,$$

the coefficients of which possess the symmetrical property

$$\Gamma^\mu_{\nu\rho} = \Gamma^\mu_{\rho\nu} \qquad . \quad . \quad . \quad (5)$$

If two vectors ξ^μ, η^μ at P are transformed by parallel displacement at P′ into $\xi^\mu + d\xi^\mu$, $\eta^\mu + d\eta^\mu$, then the postulate of similarity stated under 1 above, which goes beyond affinity, tells us that

$$\sum_{\mu\nu}(g_{\mu\nu} + dg_{\mu\nu})(\xi^\mu + d\xi^\mu)(\eta^\nu + d\eta^\nu)$$

must be proportional to

$$\sum_{\mu\nu} g_{\mu\nu}\xi^\mu\eta^\nu.$$

If we call the factor of proportionality, which differs infinitesimally from 1, $1 + d\phi$, and define the reduction of an index in the usual way by the formula

$$a_\mu = \sum_\nu g_{\mu\nu}a^\nu,$$

we obtain

$$dg_{\mu\nu} - (d\gamma_{\nu\mu} + d\gamma_{\mu\nu}) = g_{\mu\nu}d\phi \quad . \quad . \quad (6)$$

From this it follows that $d\phi$ is a linear differential form

$$d\phi = \sum_\mu \phi_\mu dx_\mu \quad . \quad . \quad . \quad . \quad (7)$$

If this is known, the equation (6) or

$$\Gamma_{\mu,\nu\rho} + \Gamma_{\nu,\mu\rho} = \frac{\partial g_{\mu\nu}}{\partial x_\rho} - g_{\mu\nu}\phi_\rho,$$

together with the condition for symmetry (5), gives unequivocally the quantities Γ. *The internal metrical connexion of space thus depends on a linear form* (7) *besides the quadratic form* (2)—which is determined except as to an arbitrary factor of proportionality.* If we substitute $\lambda g_{\mu\nu}$ for $g_{\mu\nu}$ with-

* [I have now modified this structure in the following points (cf. the final presentation in ed. 4 of "Raum, Zeit, Materie," 1921, §§ 13, 18). (*a*) In place of postulates 1 and 2, which the parallel displacement has to fulfil, there is now one postulate: Let there be a system of co ordinates at the point P, by the employment of which the components of every vector at P are not altered by parallel displacement to any point in infinite proximity to P. This postulate characterizes the essence of the parallel displacement as that of a transposition, concerning which it may be correctly asserted that it leaves the vectors "unaltered." (*b*) To the metrics at the single point P, according to which there is attached to every vector $\mathbf{x} = \xi^\mu$ at P a tract of such a kind that two vectors define the same tract when, and only when, they possess the same measure-number $l = \Sigma g_{\mu\nu}\xi^\mu\xi^\nu$, there must now be added the metrical connexion of P with the points in its neighbourhood: by congruent transposition to the infinitely near point P′ a tract at P passes over into a definite tract at P′. If we make a requirement of this concept of congruent transposition of tracts analogous to that which has just been postulated, under (*a*), of the concept of parallel displacement of vectors, we see that this process (in which the measure-number l of the tract is increased by dl) is expressed in the equations

$$dl = l d\phi\,;\; d\phi = \Sigma\phi_\mu dx_\mu.$$

out changing the system of co-ordinates, the quantities $d\gamma^{\mu}_{\nu}$ do not change, $d\gamma_{\mu\nu}$ assumes the factor λ, and $dg_{\mu\nu}$ becomes $\lambda dg_{\mu\nu} + g_{\mu\nu}d\lambda$. Equation (6) then shows that $d\phi$ becomes

$$d\phi + \frac{d\lambda}{\lambda} = d\phi + d\,(\log\lambda).$$

What remains undetermined, therefore, in the linear form $\Sigma\phi_{\mu}dx_{\mu}$ is not a factor of proportionality which would have to be settled by an arbitrary choice of a unit of measurement, but, rather, the arbitrary element inherent in it consists in an additive total differential. For the analytical representation of geometry the forms

$$g_{\mu\nu}dx_{\mu}dx_{\nu},\ \ \phi_{\mu}dx_{\mu}\ . \qquad . \quad . \quad . \quad (8)$$

are on an equal footing with

$$\lambda \cdot g_{\mu\nu}dx_{\mu}dx_{\nu} \text{ and } \phi_{\mu}dx_{\mu} + d\,(\log\lambda) \quad . \quad . \quad (9)$$

where λ is any positive function of position. Hence there is invariant significance in the anti-symmetrical tensor with the components

$$\mathrm{F}_{\mu\nu} = \frac{\partial\phi_{\mu}}{\partial x_{\nu}} - \frac{\partial\phi_{\nu}}{\partial x_{\mu}} \quad . \quad . \quad . \quad (10)$$

i.e. the form

$$\mathrm{F}_{\mu\nu} = dx_{\mu}\delta x_{\nu} = \tfrac{1}{2}\mathrm{F}_{\mu\nu}\Delta x_{\mu\nu}$$

which depends bilinearly on two arbitrary displacements dx and δx at the point P—or, rather, depends linearly on the surface element with the components $\Delta x_{\mu\nu} = dx_{\mu}\delta x_{\nu} - dx_{\nu}\delta x_{\mu}$ which is defined by these two displacements. The special case of the theory as hitherto developed, in which the arbitrarily chosen unit of length at the origin allows itself to be transferred by parallel displacement to all points of space in a manner which is independent of the path traversed —this special case occurs when the $g_{\mu\nu}$ can be absolutely determined in such a way that the ϕ_{μ} vanish. The $\Gamma^{\mu}_{\nu\rho}$ are

In these circumstances the metrics and the metrical connexion determine the "affine" connexion (parallel displacement) without ambiguity—and indeed, according to my present view of the problem of space this is the most fundamental fact of geometry—whereas according to the presentation given in the text it is the linear form $d\phi$ that remains arbitrary in the given metrics at the parallel displacement.]

then nothing else than the Christoffel three-indices symbols. The necessary and sufficient invariant condition for the occurrence of this case consists in the identical vanishing of the tensor $F_{\mu\nu}$.

This naturally suggests interpreting ϕ_μ in world-geometry as the four-potential, and the tensor F consequently as electromagnetic field. For the absence of an electromagnetic field is the necessary condition for the validity of Einstein's theory, which, up to the present, accounts for the phenomena of gravitation only. If this view is accepted, it will be seen that the electric quantities are of such a nature that their characterization by numbers in a definite system of co-ordinates does not depend on the arbitrary choice of a unit of measurement. In fact, in the question of the unit of measurement and of dimension there must be a new orientation of the theory. Hitherto a quantity has been spoken of as, e.g., a tensor of the second rank, if a single value of the quantity determines a matrix of numbers $a_{\mu\nu}$ in each system of co-ordinates after an arbitrary unit of measurement has been selected, these numbers forming the coefficients of an invariant bilinear form of two arbitrary, infinitesimal displacements

$$a_{\mu\nu}dx_\mu \delta x_\nu \quad . \quad . \quad . \quad . \quad (11)$$

But here we speak of a tensor, if, with a system of co-ordinates taken as a base, and after definite selection of the factor of proportionality contained in the $g_{\mu\nu}$, the components $a_{\mu\nu}$ are determined without ambiguity and in such a way that on transforming the co-ordinates the form (11) remains invariant, but on replacing $g_{\mu\nu}$ by $\lambda g_{\mu\nu}$ the $a_{\mu\nu}$ become $\lambda^e a_{\mu\nu}$. We then say that the tensor has the weight e, or, ascribing to the linear element ds the dimension "length $= l$," that it is of dimension l^{2e}. Only those tensors of weight 0 are absolutely invariant. The field tensor with the components $F_{\mu\nu}$ is of this kind. By (10) it satisfies the first system of the Maxwell equations

$$\frac{\partial F_{\nu\rho}}{\partial x_\mu} + \frac{\partial F_{\rho\mu}}{\partial x_\nu} + \frac{\partial F_{\mu\nu}}{\partial x_\rho} = 0.$$

When once the idea of parallel displacement is clear, geometry and the tensor calculus can be established without difficulty.

(*a*) *Geodetic Lines.*—Given a point P and at that point a vector, the geodetic line from P in the direction of this vector is given by continuously moving the vector parallel to itself in its own direction. Employing a suitable parameter τ the differential equation of the geodetic line is

$$\frac{d^2 x_\mu}{d\tau^2} + \Gamma^\mu_{\nu\rho} \frac{dx_\nu}{d\tau} \frac{dx_\rho}{d\tau} = 0.$$

(Of course it cannot be characterized as the line of smallest length, because the notion of curve-length has no meaning.)

(*b*) *Tensor Calculus.*—To deduce, for example, a tensor field of rank 2 by differentiation from a covariant tensor field of rank 1 and weight 0 with components f_μ, we call in the help of an arbitrary vector ξ^μ at the point P, form the invariant $f_\mu \xi^\mu$ and its infinitely small alteration on transition from the point P with the co-ordinates x_μ to the neighbouring point P′ with the co-ordinates $x_\mu + dx_\mu$ by shifting the vector along a parallel to itself during this transition. For this alteration we have

$$\frac{\partial f_\mu}{\partial x_\nu} \xi^\mu dx_\nu + f_\rho d\xi^\rho = \left(\frac{\partial f_\mu}{\partial x_\nu} - \Gamma^\rho_{\mu\nu} f_\rho \right) \xi^\mu dx_\nu.$$

The quantities in brackets on the right are therefore the components of a tensor field of rank 2 and weight 0, which is formed from the field f in a perfectly invariant manner.

(*c*) *Curvature.*—To construct the analogue to Riemann's tensor of curvature, let us begin with the figure employed above, of an infinitely small parallelogram, consisting of the points P, P_1, P_2, and $P_{12} = P_{21}$.* If we displace a vector $\mathbf{x} = \xi^\mu$ at P parallel to itself, to P_1 and from there to P_{12}, and a second time first to P_2 and thence to P_{21}, then, since P_{12} and P_{21} coincide, there is a meaning in forming the difference $\Delta\mathbf{x}$ of the two vectors obtained at this point. For their components we have

$$\Delta\xi^\mu = \Delta R^\mu_\nu \xi^\nu \quad . \quad . \quad . \quad . \quad (12)$$

where the ΔR^μ_ν are independent of the displaced vector $\mathbf{x}$, but

* [Here it is not essential that opposite sides of the infinitely small "parallelogram" are produced by parallel displacement one from the other; we are concerned only with the coincidence of the points P_{12} and P_{21}.]

on the other hand depend linearly on the surface-element defined by the two displacements $PP_1 = dx_\mu$, $PP_2 = \delta x_\mu$. Thus

$$\Delta R^\mu_\nu = R^\mu_{\nu\rho\sigma} dx_\rho \delta x_\sigma = \tfrac{1}{2} R^\mu_{\nu\rho\sigma} \Delta x_{\rho\sigma}.$$

The components of curvature $R^\mu_{\nu\rho\sigma}$, depending solely on the place P, possess the two properties of symmetry that (1) they change sign on the interchange of the last two indices ρ and σ, and (2), if we perform the three cylic interchanges $\nu\rho\sigma$, and add up the appropriate components, the result is 0. Reducing the index μ, we obtain at $R_{\mu\nu\rho\sigma}$ the components of a covariant tensor of rank 4 and weight 1. Even without calculation we see that R divides in a natural, invariant manner into two parts,

$$R^\mu_{\nu\rho\sigma} = P^\mu_{\nu\rho\sigma} - \tfrac{1}{2}\delta^\mu_\nu F_{\rho\sigma} \quad (\delta^\mu_\nu = 1 \text{ if } \mu = \nu : \ = 0 \text{ if } \mu \neq \nu), \quad (13)$$

of which the first, $P^\mu_{\nu\rho\sigma}$, is anti-symmetrical, not only in the indices $\rho\sigma$, but also in μ and ν. Whereas the equations $F_{\mu\nu} = 0$ characterize our space as one without an electromagnetic field, i.e. as one in which the problem of the conveyance of length is integrable, the equations $P^\mu_{\nu\rho\sigma} = 0$ are, as (13) shows, the invariant conditions for the absence of a gravitational field, i.e. for the problem of the conveyance of direction to be integrable. The Euclidean space alone is one which at the same time is free of electricity and of gravitation.

The simplest invariant of a linear copy like (12), which to each vector **x** assigns a vector Δ**x**, is its "spur"

$$\frac{1}{n}\Delta R^\mu_\mu.$$

For this, by (13), we obtain in the present case the form

$$-\tfrac{1}{2}F_{\rho\sigma}dx_\rho\delta x_\sigma$$

which we have already encountered above. The simplest invariant of a tensor like $-\tfrac{1}{2}F_{\rho\sigma}$ is the "square of its magnitude"

$$L = \tfrac{1}{4}F_{\rho\sigma}F^{\rho\sigma} \quad . \quad . \quad . \quad . \quad (14)$$

L is evidently an invariant of weight -2, because the tensor F has weight 0. If g is the negative determinant of the $g_{\mu\nu}$, and

$$d\omega = \sqrt{g}\,dx_0 dx_1 dx_2 dx_3 = \sqrt{g}\,dx$$

the volume of an infinitely small element of volume, it is known that the Maxwell theory is governed by the quantity of electrical action, which is equal to the integral $\int L d\omega$ of this simplest invariant, extended over any chosen territory, and indeed is governed in the sense that, with any variations of the $g_{\mu\nu}$ and ϕ_μ, which vanish at the limits of world-territory, we have

$$\delta\int L d\omega = \int (S^\mu d\phi_\mu + T^{\mu\nu}\delta g_{\mu\nu})d\omega,$$

where

$$S^\mu = \frac{1}{\sqrt{g}} \frac{\partial(\sqrt{g}F^{\mu\nu})}{\partial x_\nu}$$

are the left-hand sides of the generalized Maxwellian equations (the right-hand sides of which are the components of the four-current), and the $T^{\mu\nu}$ form the energy-momentum tensor of the electromagnetic field. As L is an invariant of weight -2, whereas the volume-element in n-dimensional geometry is an invariant of weight $\frac{1}{2}n$, the integral has significance only when the number of dimensions $n = 4$. Thus on our interpretation the possibility of the Maxwell theory is restricted to the case of four dimensions. In the four-dimensional world, however, the quantity of electromagnetic action becomes a pure number. Nevertheless, the magnitude of the quantity 1 cannot be ascertained in the traditional units of the c.g.s. system until a physical problem, to be tested by observation (as for example the electron), has been calculated on the basis of our theory.

Passing now from geometry to physics, we have to assume, following the precedent of Mie's theory,* that all the laws of nature rest upon a definite integral invariant, the action-quantity

$$\int W d\omega = \int \mathfrak{W} dx, \quad \mathfrak{W} = W\sqrt{g},$$

in such a way that the real world is distinguished from all other possible four-dimensional metrical spaces by the characteristic that for it the action-quantity contained in any part of its domain assumes a stationary value in relation to such variations of the potentials $g_{\mu\nu}$, ϕ_μ as vanish at the limits of

* Ann. d. Physik, 37, 39, 40, 1912-13.

the territory in question. W, the world-density of the action, must be an invariant of weight − 2. The action-quantity is in any case a pure number; thus our theory at once accounts for that atomistic structure of the world to which current views attach the most fundamental importance—the action-quantum. The simplest and most natural conjecture which we can make for W, is

$$W = R^{\mu}_{\nu\rho\sigma} R^{\nu\rho\sigma}_{\mu} = | R |^2.$$

For this we also have, by (13),

$$W = | P |^2 + 4L.$$

(There could be no doubt about anything here except perhaps the factor 4, with which the electric term L is added to the first.) But even without particularizing the action-quantity we can draw some general conclusions from the principle of action. For we shall show that as, according to investigations by Hilbert, Lorentz, Einstein, Klein, and the author,* the four laws of the conservation of matter (the energy-momentum tensor) are connected with the invariance of the action quantity (containing four arbitrary functions) with respect to transformations of co-ordinates, so in the same way the law of the conservation of electricity is connected with the "measure-invariance" [transition from (8) to (9)] which here makes its appearance for the first time, introducing a fifth arbitrary function. The manner in which the latter associates itself with the principles of energy and momentum seems to me one of the strongest general arguments in favour of the theory here set out—so far as there can be any question at all of confirmation in purely speculative matters.

For any variation which vanishes at the limits of the world-territory under consideration we have

$$\delta \int \mathfrak{W} dx = \int (\mathfrak{W}^{\mu\nu} \delta g_{\mu\nu} + \mathfrak{w}^{\mu} \delta \phi_{\mu}) dx \quad (\mathfrak{W}^{\mu\nu} = \mathfrak{W}^{\mu\nu}) \quad (15)$$

* Hilbert, "Die Grundlagen der Physik," Göttinger Nachrichten, 20 Nov., 1915; H. A. Lorentz in four papers in the Versl. K. Ak. van Wetensch., Amsterdam, 1915-16; A. Einstein, Berl. Ber., 1916, pp. 1111-6; F. Klein, Gott. Nachr., 25 Jan., 1918; H. Weyl, Ann. d. Physik, 54, 1917, pp. 121-5.

The laws of nature then take the form

$$\mathfrak{W}^{\mu\nu} = 0, \mathfrak{w}^{\mu} = 0 \quad . \quad . \quad . \quad (16)$$

The former may be regarded as the laws of the gravitational field, the latter as those of the electromagnetic field. The quantities W_{ν}^{μ}, w^{μ} defined by

$$\mathfrak{W}_{\nu}^{\mu} = \sqrt{g} W_{\nu}^{\mu}, \quad \mathfrak{w}^{\mu} = \sqrt{g} w^{\mu}$$

are the mixed or, respectively, the contravariant components of a tensor of rank 2 or 1 respectively, and of weight -2. In the system of equations (16) there are five which are redundant, in accordance with the properties of invariance. This is expressed in the following five invariant identities, which subsist between their left-hand sides :—

$$\frac{\partial \mathfrak{w}^{\mu}}{\partial x_{\mu}} \equiv \mathfrak{W}_{\mu}^{\mu} \quad . \quad . \quad . \quad . \quad (17)$$

$$\frac{\partial \mathfrak{W}_{\nu}^{\mu}}{\partial x_{\mu}} - \Gamma_{\nu\beta}^{\alpha} \mathfrak{W}_{\alpha}^{\beta} \equiv \tfrac{1}{2} F_{\mu\nu} \mathfrak{w}^{\mu} \quad . \quad . \quad . \quad (18)$$

The first results from the measure-invariance. For if in the transition from (8) to (9) we assume for log λ an infinitely small function of position $\delta\rho$, we obtain the variation

$$\delta g_{\mu\nu} = g_{\mu\nu}\delta\rho, \quad \delta\phi_{\mu} = \frac{\partial(\delta\rho)}{\partial x_{\mu}}.$$

For this variation (15) must vanish. In the second place if we utilize the invariance of the action-quantity with respect to transformations of co-ordinates by means of an infinitely small deformation of the world - continuum,* we obtain the identities

$$\frac{\partial \mathfrak{W}_{\nu}^{\mu}}{\partial x_{\mu}} - \tfrac{1}{2} \frac{\partial g_{\alpha\beta}}{\partial x_{\nu}} \mathfrak{W}^{\alpha\beta} + \tfrac{1}{2}\left(\frac{\partial \mathfrak{w}^{\mu}}{\partial x_{\mu}} \phi_{\nu} - \Gamma_{\alpha\nu} \mathfrak{w}^{\alpha}\right) \equiv 0$$

which change into (18) when, by (17) $\partial \mathfrak{w}^{\mu}/\partial x_{\mu}$ is replaced by $g_{\alpha\beta}\mathfrak{W}^{\alpha\beta}$

From the gravitational laws alone therefore we already obtain

$$\frac{\partial \mathfrak{w}^{\mu}}{\partial x_{\mu}} = 0, \quad . \quad . \quad . \quad . \quad . \quad (19)$$

* Weyl, Ann. d. Physik, 54, 1917, pp. 121-5; F. Klein, Gött. Nachr., 25 Jan., 1918.

and from the laws of the electromagnetic field alone

$$\frac{\partial}{\partial x_\mu}\mathfrak{W}^\mu_\nu - \Gamma^\alpha_{\nu\beta}\mathfrak{W}^\beta_\alpha = 0 \quad . \quad . \quad . \quad (20)$$

In Maxwell's theory $\mathfrak{w}^\mu$ has the form

$$\mathfrak{w}^\mu \equiv \frac{\partial(\sqrt{g}F^{\mu\nu})}{\partial x_\nu} - \mathfrak{s}^\mu, \quad \mathfrak{s}^\mu = \sqrt{g}s^\mu$$

where s^μ denotes the four-current. Since the first part here satisfies the equation (19) identically, this equation gives us the law of conservation of electricity

$$\frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g}s^\nu)}{\partial x_\mu} = 0.$$

Similarly in Einstein's theory of gravitation $\mathfrak{W}^\mu_\nu$ consists of two terms, the first of which satisfies equation (20) identically, and the second is equal to the mixed components of the energy-momentum tensor T^μ_ν multiplied by $\sqrt{g}$. Thus equations (20) lead to the four laws of the conservation of matter. Quite analogous circumstances hold good in our theory if we choose the form (14) for the action-quantity. The five principles of conservation are "eliminants" of the field laws, i.e. they follow from them in a twofold manner, and thus demonstrate that among them there are five which are redundant.

With the form (14) for the action-quantity the Maxwell equations run, for example :—

$$\frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g}F^{\mu\nu})}{\partial x_\nu} = s^\mu, \quad . \quad . \quad . \quad (21)$$

and the current is

$$s_\mu = \tfrac{1}{4}\left(R\phi_\mu + \frac{\partial R}{\partial x_\mu}\right),$$

where R denotes that invariant of weight -1 which arises from $R^\mu_{\nu\rho\sigma}$ if we first contract with respect to μ, ρ and then with respect to ν and σ. If R^* denotes Riemann's invariant of curvature constructed solely from the $g_{\mu\nu}$, calculation gives

$$R = R^* - \frac{3}{\sqrt{g}} \frac{\partial(\sqrt{g}\phi^\mu)}{\partial x_\mu} + \frac{3}{2}\phi_\mu\phi^\mu.$$

In the static case, where the space components of the electromagnetic potential disappear, and all quantities are independent of the time x_0, by (21) we must have

$$R = R^* + \frac{3}{2}\phi_0\phi^0 = \text{const.}$$

But in a world-territory in which $R \neq 0$ we may make $R = \text{const.} = \pm 1$ everywhere, by appropriate determination of the unit of length. Only we have to expect, under conditions which are variable with time, surfaces $R = 0$, which evidently will play some singular part. R cannot be used as density of action (represented by R* in Einstein's theory of gravitation) because it has not the weight − 2. The consequence is that though our theory leads to Maxwell's electromagnetic equations, it does not lead to Einstein's gravitation equations. In their place appear differential equations of order 4. But indeed it is very improbable that Einstein's equations of gravitation are strictly correct, because, above all things, the gravitation constant occurring in them is not at all in the picture with the other constants of nature, the gravitation radius of the charge and mass of an electron, for example, being of an entirely different order of magnitude (10^{20} or 10^{40} times as small) from that of the radius of the electron itself.*

It was my intention here merely to develop briefly the general principles of the theory.† The problem naturally

* Cf. Weyl, Ann. d. Physik, 54, 1917, p. 133.

† [The problem of defining all W invariants allowable as action-quantities, under the requirement that they should contain the derivatives of the $g_{\mu\nu}$ only to the second order at most, and those of the ϕ_μ only to the first order, was solved by R. Weitzenbock (Sitzungsber. d. Akad. d. Wissensch. in Wien, 129, 1920; 130, 1921). If we omit the invariants W for which the variation $\delta\int W d\omega$ vanishes identically, there remain according to a later calculation by R. Bach (Math. Zeitschrift, 9, 1921, pp. 125 and 169) only three combinations. The real W seems to be a linear combination of Maxwell's L and the square of R. This conjecture has been tested more carefully by W. Pauli (Physik. Zeitschrift., 20, 1919 pp. 457-67) and myself: in particular we succeeded in advancing so far on this basis as to deduce the equations of

presents itself of deducing the physical consequences of the theory on the basis of the special form for the action-quantity given in (14), and of comparing these with experience, examining particularly whether the existence of the electron and the peculiarities of the hitherto unexplained processes in the atom can be deduced from the theory.* The task is extraordinarily complicated from the mathematical point of view, because it is impossible to obtain approximate solutions if we restrict ourselves to the linear terms; for since it is certainly not permissible to neglect terms of higher order in the interior of the electron, the linear equations obtained by neglecting these may have, in general, only the solution 0. I propose to return to all these matters in greater detail in another place.

motion of a material particle. The invariant (14) selected above, at hazard in the first place, seems on the contrary to play no part in nature. Cf. Raum, "Zeit, Materie," ed. 4, §§ 35, 36, or Weyl, Physik. Zeitschr., 22, 1921, pp. 473-80.]

* [Meanwhile I have quite abandoned these hopes, raised by Mie's theory; I do not believe that the problem of matter is to be solved by a mere field theory. Cf. on this subject my article "Feld und Materie," Ann. d. Physik, 65, 1921, pp. 541-63.]

ESSAY I*

ON THE SPECTRUM OF HYDROGEN

Empirical spectral laws. Hydrogen possesses not only the smallest atomic weight of all the elements, but it also occupies a peculiar position both with regard to its physical and its chemical properties. One of the points where this becomes particularly apparent is the hydrogen line spectrum.

The spectrum of hydrogen observed in an ordinary Geissler tube consists of a series of lines, the strongest of which lies at the red end of the spectrum, while the others extend out into the ultra violet, the distance between the various lines, as well as their intensities, constantly decreasing. In the ultra violet the series converges to a limit.

Balmer, as we know, discovered (1885) that it was possible to represent the wave lengths of these lines very accurately by the simple law

$$\frac{1}{\lambda_n} = R\left(\frac{1}{4} - \frac{1}{n^2}\right), \qquad \text{......................(1)}$$

where R is a constant and n is a whole number. The wave lengths of the five strongest hydrogen lines, corresponding to $n = 3, 4, 5, 6, 7$, measured in air at ordinary pressure and temperature, and the values of these wave lengths multiplied by $\left(\frac{1}{4} - \frac{1}{n^2}\right)$ are given in the following table:

n	$\lambda . 10^8$	$\lambda . \left(\frac{1}{4} - \frac{1}{n^2}\right) . 10^{10}$
3	6563·04	91153·3
4	4861·49	91152·9
5	4340·66	91153·9
6	4101·85	91152·2
7	3970·25	91153·7

The table shows that the product is nearly constant, while the deviations are not greater than might be ascribed to experimental errors.

As you already know, Balmer's discovery of the law relating to the hydrogen spectrum led to the discovery of laws applying to the spectra of other elements. The most important work in this

* Address delivered before the Physical Society in Copenhagen, Dec. 20, 1913.

connection was done by Rydberg (1890) and Ritz (1908). Rydberg pointed out that the spectra of many elements contain series of lines whose wave lengths are given approximately by the formula

$$\frac{1}{\lambda_n} = A - \frac{R}{(n+\alpha)^2},$$

where A and α are constants having different values for the various series, while R is a universal constant equal to the constant in the spectrum of hydrogen. If the wave lengths are measured in vacuo Rydberg calculated the value of R to be 109675. In the spectra of many elements, as opposed to the simple spectrum of hydrogen, there are several series of lines whose wave lengths are to a close approximation given by Rydberg's formula if different values are assigned to the constants A and α. Rydberg showed, however, in his earliest work, that certain relations existed between the constants in the various series of the spectrum of one and the same element. These relations were later very successfully generalized by Ritz through the establishment of the "combination principle." According to this principle, the wave lengths of the various lines in the spectrum of an element may be expressed by the formula

$$\frac{1}{\lambda} = F_r(n_1) - F_s(n_2). \quad \text{.....................(2)}$$

In this formula n_1 and n_2 are whole numbers, and $F_1(n), F_2(n), \ldots$ is a series of functions of n, which may be written approximately

$$F_r(n) = \frac{R}{(n+\alpha_r)^2},$$

where R is Rydberg's universal constant and α_r is a constant which is different for the different functions. A particular spectral line will, according to this principle, correspond to each combination of n_1 and n_2, as well as to the functions $F_1, F_2, \ldots$. The establishment of this principle led therefore to the prediction of a great number of lines which were not included in the spectral formulae previously considered, and in a large number of cases the calculations were found to be in close agreement with the experimental observations. In the case of hydrogen Ritz assumed that formula (1) was a special case of the general formula

$$\frac{1}{\lambda} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right), \quad \text{........................(3)}$$

and therefore predicted among other things a series of lines in the infra red given by the formula

$$\frac{1}{\lambda}=R\left(\frac{1}{9}-\frac{1}{n^2}\right).$$

In 1909 Paschen succeeded in observing the first two lines of this series corresponding to $n=4$ and $n=5$.

The part played by hydrogen in the development of our knowledge of the spectral laws is not solely due to its ordinary simple spectrum, but it can also be traced in other less direct ways. At a time when Rydberg's laws were still in want of further confirmation Pickering (1897) found in the spectrum of a star a series of lines whose wave lengths showed a very simple relation to the ordinary hydrogen spectrum, since to a very close approximation they could be expressed by the formula

$$\frac{1}{\lambda}=R\left(\frac{1}{4}-\frac{1}{(n+\frac{1}{2})^2}\right).$$

Rydberg considered these lines to represent a new series of lines in the spectrum of hydrogen, and predicted according to his theory the existence of still another series of hydrogen lines the wave lengths of which would be given by

$$\frac{1}{\lambda}=R\left(\frac{1}{(\frac{3}{2})^2}-\frac{1}{n^2}\right).$$

By examining earlier observations it was actually found that a line had been observed in the spectrum of certain stars which coincided closely with the first line in this series (corresponding to $n=2$); from analogy with other spectra it was also to be expected that this would be the strongest line. This was regarded as a great triumph for Rydberg's theory and tended to remove all doubt that the new spectrum was actually due to hydrogen. Rydberg's view has therefore been generally accepted by physicists up to the present moment. Recently however the question has been reopened and Fowler (1912) has succeeded in observing the Pickering lines in ordinary laboratory experiments. We shall return to this question again later.

The discovery of these beautiful and simple laws concerning the line spectra of the elements has naturally resulted in many attempts at a theoretical explanation. Such attempts are very alluring because the simplicity of the spectral laws and the exceptional accuracy with which they apply appear to promise that the correct explanation will be very simple and will give valuable information about the properties of matter. I should like to consider some of these theories somewhat more closely, several of which are extremely interesting and have been developed with the greatest keenness and ingenuity, but unfortunately space does not permit me to do so here. I shall have to limit myself to the statement that not one of the theories so far proposed appears to offer a satisfactory or even a plausible way of explaining the laws of the line spectra. Considering our deficient knowledge of the laws which determine the processes inside atoms it is scarcely possible to give an explanation of the kind attempted in these theories. The inadequacy of our ordinary theoretical conceptions has become especially apparent from the important results which have been obtained in recent years from the theoretical and experimental study of the laws of temperature radiation. You will therefore understand that I shall not attempt to propose an explanation of the spectral laws; on the contrary I shall try to indicate a way in which it appears possible to bring the spectral laws into close connection with other properties of the elements, which appear to be equally inexplicable on the basis of the present state of the science. In these considerations I shall employ the results obtained from the study of temperature radiation as well as the view of atomic structure which has been reached by the study of the radioactive elements.

Laws of temperature radiation. I shall commence by mentioning the conclusions which have been drawn from experimental and theoretical work on temperature radiation.

Let us consider an enclosure surrounded by bodies which are in temperature equilibrium. In this space there will be a certain amount of energy contained in the rays emitted by the surrounding substances and crossing each other in every direction. By making the assumption that the temperature equilibrium will not be disturbed by the mutual radiation of the various bodies Kirchhoff (1860) showed that the amount of energy per unit volume as well as the distribution of this energy among the various wave lengths is independent of the form and size of the space and of the nature of the surrounding bodies and depends only on the temperature. Kirchhoff's result has been confirmed by experiment, and the amount of energy and its distribution among the various wave lengths and the manner in which it depends on the temperature are now fairly well known from a great amount of experimental work; or, as it is usually expressed, we have a fairly accurate experimental knowledge of the "laws of temperature radiation."

Kirchhoff's considerations were only capable of predicting the existence of a law of temperature radiation, and many physicists have subsequently attempted to find a more thorough explanation of the experimental results. You will perceive that the electromagnetic theory of light together with the electron theory suggests a method of solving this problem. According to the electron theory of matter a body consists of a system of electrons. By making certain definite assumptions concerning the forces acting on the electrons it is possible to calculate their motion and consequently the energy radiated from the body per second in the form of electromagnetic oscillations of various wave lengths. In a similar manner the absorption of rays of a given wave length by a substance can be determined by calculating the effect of electromagnetic oscillations upon the motion of the electrons. Having investigated the emission and absorption of a body at all temperatures, and for rays of all wave lengths, it is possible, as Kirchhoff has shown, to determine immediately the laws of temperature radiation. Since the result is to be independent of the nature of the body we are justified in expecting an agreement with experiment, even though very special assumptions are made about the forces acting upon the electrons of the hypothetical substance. This naturally simplifies the problem considerably, but it is nevertheless sufficiently difficult and it is remarkable that it has been possible to make any advance at all in this direction. As is well known this has been done by Lorentz (1903). He calculated the emissive as well as the absorptive power of a metal for long wave lengths, using the same assumptions about the motions of the electrons in the metal that Drude (1900) employed in his calculation of the ratio of the electrical and thermal conductivities. Subsequently, by calculating the ratio of the emissive to the absorptive power, Lorentz really obtained an expression for the law of temperature radiation which for long wave lengths agrees remarkably well with experimental facts. In spite of this beautiful and promising result, it has nevertheless become apparent that the electromagnetic theory is incapable of explaining the law of temperature radiation. For, it is possible to show, that, if the investigation is not confined to oscillations of long wave lengths, as in Lorentz's work, but is also extended to oscillations corresponding to small wave lengths, results are obtained which are contrary to experiment. This is especially evident from Jeans' investigations (1905) in which he employed a very interesting statistical method first proposed by Lord Rayleigh.

We are therefore compelled to assume, that the classical electrodynamics does not agree with reality, or expressed more carefully, that it can not be employed in calculating the absorption and emission of radiation by atoms. Fortunately, the law of temperature radiation has also successfully indicated the direction in which the necessary changes in the electrodynamics are to be sought. Even before the appearance of the papers by Lorentz and Jeans, Planck (1900) had derived theoretically a formula for the black body radiation which was in good agreement with the results of experiment. Planck did not limit himself exclusively to the classical electrodynamics, but introduced the further assumption that a system of oscillating electrical particles (elementary resonators) will neither radiate nor absorb energy continuously, as required by the ordinary electrodynamics, but on the contrary will radiate and absorb discontinuously. The energy contained within the system at any moment is always equal to a whole multiple of the so-called quantum of energy the magnitude of which is equal to $h\nu$, where h is Planck's constant and ν is the frequency of oscillation of the system per second. In formal respects Planck's theory leaves much to be desired; in certain calculations the ordinary electrodynamics is used, while in others assumptions distinctly at variance with it are introduced without any attempt being made to show that it is possible to give a consistent explanation of the procedure used. Planck's theory would hardly have acquired general recognition merely on the ground of its agreement with experiments on black body radiation, but, as you know, the theory has also contributed quite remarkably to the elucidation of many different physical phenomena, such as specific heats, photoelectric effect, X-rays and the absorption of heat rays by gases. These explanations involve more than the qualitative assumption of a discontinuous transformation of energy, for with the aid of Planck's constant h it seems to be possible, at least approximately, to account for a great number of phenomena about which nothing could be said previously. It is therefore hardly too early to express the opinion that, whatever the final explanation will be, the discovery of "energy quanta" must be considered as one of the most important results arrived at in physics, and must be taken into consideration in investigations of the properties of atoms and particularly in connection with any explanation of the spectral laws in which such phenomena as the emission and absorption of electromagnetic radiation are concerned.

The nuclear theory of the atom. We shall now consider the second part of the foundation on which we shall build, namely the conclusions arrived at from experiments with the rays emitted by radioactive substances. I have previously here in the Physical Society had the opportunity of speaking of the scattering of α rays in passing through thin plates, and to mention how Rutherford (1911) has proposed a theory for the structure of the atom in order to explain the remarkable and unexpected results of these experiments. I shall, therefore, only remind you that the characteristic feature of Rutherford's theory is the assumption of the existence of a positively charged nucleus inside the atom. A number of electrons are supposed to revolve in closed orbits around the nucleus, the number of these electrons being sufficient to neutralize the positive charge of the nucleus. The dimensions of the nucleus are supposed to be very small in comparison with the dimensions of the orbits of the electrons, and almost the entire mass of the atom is supposed to be concentrated in the nucleus.

According to Rutherford's calculation the positive charge of the nucleus corresponds to a number of electrons equal to about half the atomic weight. This number coincides approximately with the number of the particular element in the periodic system and it is therefore natural to assume that the number of electrons in the atom is exactly equal to this number. This hypothesis, which was first stated by van den Broek (1912), opens the possibility of obtaining a simple explanation of the periodic system. This assumption is strongly confirmed by experiments on the elements of small atomic weight. In the first place, it is evident that according to Rutherford's theory the α particle is the same as the nucleus of a helium atom. Since the α particle has a double positive charge it follows immediately that a neutral helium atom contains two electrons. Further the concordant results obtained from calculations based on experiments as different as the diffuse scattering of X-rays and the decrease in velocity of α rays in passing through matter render the conclusion extremely likely that a hydrogen atom contains only a single electron. This agrees most beautifully with the fact that J. J. Thomson in his well-known experiments on rays of positive electricity has never observed a hydrogen atom with more than a single positive charge, while all other elements investigated may have several charges.

Let us now assume that a hydrogen atom simply consists of an electron revolving around a nucleus of equal and opposite charge, and of a mass which is very large in comparison with that of the electron. It is evident that this assumption may explain the peculiar position already referred to which hydrogen occupies among the elements, but it appears at the outset completely hopeless to attempt to explain anything at all of the special properties of hydrogen, still less its line spectrum, on the basis of considerations relating to such a simple system.

Let us assume for the sake of brevity that the mass of the nucleus is infinitely large in proportion to that of the electron, and that the velocity of the electron is very small in comparison with that of light. If we now temporarily disregard the energy radiation, which, according to the ordinary electrodynamics, will accompany the accelerated motion of the electron, the latter in accordance with Kepler's first law will describe an ellipse with the nucleus in one of the foci. Denoting the frequency of revolution by ω, and the major axis of the ellipse by $2a$ we find that

$$\omega^2 = \frac{2W^3}{\pi^2 e^4 m}, \quad 2a = \frac{e^2}{W}, \quad \text{..................(4)}$$

where e is the charge of the electron and m its mass, while W is the work which must be added to the system in order to remove the electron to an infinite distance from the nucleus.

These expressions are extremely simple and they show that the magnitude of the frequency of revolution as well as the length of the major axis depend only on W, and are independent of the excentricity of the orbit. By varying W we may obtain all possible values for ω and $2a$. This condition shows, however, that it is not possible to employ the above formulae directly in calculating the orbit of the electron in a hydrogen atom. For this it will be necessary to assume that the orbit of the electron can not take on all values, and in any event, the line spectrum clearly indicates that the oscillations of the electron cannot vary continuously between wide limits. The impossibility of making any progress with a simple system like the one considered here might have been foretold from a consideration of the dimensions involved; for with the aid of e and m alone it is impossible to obtain a quantity which can be interpreted as a diameter of an atom or as a frequency.

If we attempt to account for the radiation of energy in the manner required by the ordinary electrodynamics it will only make matters worse. As a result of the radiation of energy W would continually increase, and the above expressions (4) show that at the same time the frequency of revolution of the system would increase, and the dimensions of the orbit decrease. This process would not stop until the particles had approached so closely to one another that they no longer attracted each other. The quantity of energy which would be radiated away before this happened would be very great. If we were to treat these particles as geometrical points this energy would be infinitely great, and with the dimensions of the electrons as calculated from their mass (about 10^{-13} cm.), and of the nucleus as calculated by Rutherford (about 10^{-12} cm.), this energy would be many times greater than the energy changes with which we are familiar in ordinary atomic processes.

It can be seen that it is impossible to employ Rutherford's atomic model so long as we confine ourselves exclusively to the ordinary electrodynamics. But this is nothing more than might have been expected. As I have mentioned we may consider it to be an established fact that it is impossible to obtain a satisfactory explanation of the experiments on temperature radiation with the aid of electrodynamics, no matter what atomic model be employed. The fact that the deficiencies of the atomic model we are considering stand out so plainly is therefore perhaps no serious drawback; even though the defects of other atomic models are much better concealed they must nevertheless be present and will be just as serious.

Quantum theory of spectra. Let us now try to overcome these difficulties by applying Planck's theory to the problem.

It is readily seen that there can be no question of a direct application of Planck's theory. This theory is concerned with the emission and absorption of energy in a system of electrical particles, which oscillate with a given frequency per second, dependent only on the nature of the system and independent of the amount of energy contained in the system. In a system consisting of an electron and a nucleus the period of oscillation corresponds to the period of revolution of the electron. But the formula (4) for ω shows that the frequency of revolution depends upon W, i.e. on the energy of the system. Still the fact that we can not immediately apply Planck's theory to our problem is not as serious as it might seem to be, for in assuming Planck's theory we have manifestly acknowledged the inadequacy of the ordinary electrodynamics and have definitely parted with the coherent group of ideas on which the latter theory is based. In fact in taking such a step we can not expect that all cases of disagreement between the theoretical conceptions hitherto employed and experiment will be removed by the use of Planck's assumption regarding the quantum of the energy momentarily present in an oscillating system. We stand here almost entirely on virgin ground, and upon introducing new assumptions we need only take care not to get into contradiction with experiment. Time will have to show to what extent this can be avoided; but the safest way is, of course, to make as few assumptions as possible.

With this in mind let us first examine the experiments on temperature radiation. The subject of direct observation is the distribution of radiant energy over oscillations of the various wave lengths. Even though we may assume that this energy comes from systems of oscillating particles, we know little or nothing about these systems. No one has ever seen a Planck's resonator, nor indeed even measured its frequency of oscillation; we can observe only the period of oscillation of the radiation which is emitted. It is therefore very convenient that it is possible to show that to obtain the laws of temperature radiation it is not necessary to make any assumptions about the systems which emit the radiation except that the amount of energy emitted each time shall be equal to $h\nu$, where h is Planck's constant and ν is the frequency of the radiation. Indeed, it is possible to derive Planck's law of radiation from this assumption alone, as shown by Debye, who employed a method which is a combination of that of Planck and of Jeans. Before considering any further the nature of the oscillating systems let us see whether it is possible to bring this assumption about the emission of radiation into agreement with the spectral laws.

If the spectrum of some element contains a spectral line corresponding to the frequency ν it will be assumed that one of the atoms of the element (or some other elementary system) can emit an amount of energy $h\nu$. Denoting the energy of the atom before and after the emission of the radiation by E_1 and E_2 we have

$$h\nu = E_1 - E_2 \text{ or } \nu = \frac{E_1}{h} - \frac{E_2}{h}. \quad \text{..............(5)}$$

During the emission of the radiation the system may be regarded as passing from one state to another; in order to introduce a name for these states, we shall call them "stationary" states, simply indicating thereby that they form some kind of waiting places between which occurs the emission of the energy corresponding to the various spectral lines. As previously mentioned the spectrum of an element consists of a series of lines whose wave lengths may be expressed by the formula (2). By comparing this expression with the relation given above it is seen that—since $\nu = \frac{c}{\lambda}$, where c is the velocity of light—each of the spectral lines may be regarded as being emitted by the transition of a system between two stationary states in which the energy apart from an additive arbitrary constant is given by $-chF_r(n_1)$ and $-chF_s(n_2)$ respectively. Using this interpretation the combination principle asserts that a series of stationary states exists for the given system, and that it can pass from one to any other of these states with the emission of a monochromatic radiation. We see, therefore, that with a simple extension of our first assumption it is possible to give a formal explanation of the most general law of line spectra.

Hydrogen spectrum. This result encourages us to make an attempt to obtain a clear conception of the stationary states which have so far only been regarded as formal. With this end in view, we naturally turn to the spectrum of hydrogen. The formula applying to this spectrum is given by the expression

$$\frac{1}{\lambda} = \frac{R}{n_1^2} - \frac{R}{n_2^2}.$$

According to our assumption this spectrum is produced by transitions between a series of stationary states of a system, concerning which we can for the present only say that the energy of the system in the nth state, apart from an additive constant, is given by $-\frac{Rhc}{n^2}$. Let us now try to find a connection between this and the model of the hydrogen atom. We assume that in the calculation of the frequency of revolution of the electron in the stationary states of the atom it will be possible to employ the above formula for ω. It is quite natural to make this assumption; since, in trying to form a reasonable conception of the stationary states, there is, for the present at least, no other means available besides the ordinary mechanics.

Corresponding to the nth stationary state in formula (4) for ω, let us by way of experiment put $W = \frac{Rhc}{n^2}$. This gives us

$$\omega_n^2 = \frac{2}{\pi^2} \frac{R^3 h^3 c^3}{e^4 m n^6}. \quad \text{......................(6)}$$

The radiation of light corresponding to a particular spectral line is according to our assumption emitted by a transition between two stationary states, corresponding to two different frequencies of revolution, and we are not justified in expecting any simple relation between these frequencies of revolution of the electron and the frequency of the emitted radiation. You understand, of course, that I am by no means trying to give what might ordinarily be described as an explanation; nothing has been said here about how or why the radiation is emitted. On one point, however, we may expect a connection with the ordinary conceptions; namely, that it will be possible to calculate the emission of slow electromagnetic oscillations on the basis of the classical electrodynamics. This assumption is very strongly supported by the result of Lorentz's calculations which have already been described. From the formula for ω it is seen that the frequency of revolution decreases as n increases, and that the expression $\frac{\omega_n}{\omega_{n+1}}$ approaches the value 1.

According to what has been said above, the frequency of the radiation corresponding to the transition between the $(n+1)$th and the nth stationary state is given by

$$\nu = Rc\left(\frac{1}{n^2} - \frac{1}{(n+1)^2}\right).$$

If n is very large this expression is approximately equal to

$$\nu = 2Rc/n^3.$$

In order to obtain a connection with the ordinary electrodynamics let us now place this frequency equal to the frequency of revolution, that is

$$\omega_n = 2Rc/n^3.$$

Introducing this value of ω_n in (6) we see that n disappears from the equation, and further that the equation will be satisfied only if

$$R = \frac{2\pi^2 e^4 m}{ch^3} \text{ ....(7)}$$

The constant R is very accurately known, and is, as I have said before, equal to 109675. By introducing the most recent values for e, m and h the expression on the right-hand side of the equation becomes equal to $1{\cdot}09 \,.\, 10^5$. The agreement is as good as could be expected, considering the uncertainty in the experimental determination of the constants e, m and h. The agreement between our calculations and the classical electrodynamics is, therefore, fully as good as we are justified in expecting.

We can not expect to obtain a corresponding explanation of the frequency values of the other stationary states. Certain simple formal relations apply, however, to all the stationary states. By introducing the expression, which has been found for R, we get for the nth state $W_n = \frac{1}{2} nh\omega_n$. This equation is entirely analogous to Planck's assumption concerning the energy of a resonator. W in our system is readily shown to be equal to the average value of the kinetic energy of the electron during a single revolution. The energy of a resonator was shown by Planck you may remember to be always equal to $nh\nu$. Further the average value of the kinetic energy of Planck's resonator is equal to its potential energy, so that the average value of the kinetic energy of the resonator, according to Planck, is equal to $\frac{1}{2}nh\omega$. This analogy suggests another manner of presenting the theory, and it was just in this way that I was originally led into these considerations. When we consider how differently the equation is employed here and in Planck's theory it appears to me misleading to use this analogy as a foundation, and in the account I have given I have tried to free myself as much as possible from it.

Let us continue with the elucidation of the calculations, and in the expression for $2a$ introduce the value of W which corresponds to the nth stationary state. This gives us

$$2a = n^2 \cdot \frac{e^2}{chR} = n^2 \cdot \frac{h^2}{2\pi^2 me^2} = n^2 \cdot 1{\cdot}1 \cdot 10^{-8}. \quad \ldots\ldots(8)$$

It is seen that for small values of n, we obtain values for the major axis of the orbit of the electron which are of the same order of magnitude as the values of the diameters of the atoms calculated from the kinetic theory of gases. For large values of n, $2a$ becomes very large in proportion to the calculated dimensions of the atoms. This, however, does not necessarily disagree with experiment. Under ordinary circumstances a hydrogen atom will probably exist only in the state corresponding to $n=1$. For this state W will have its greatest value and, consequently, the atom will have emitted the largest amount of energy possible; this will therefore represent the most stable state of the atom from which the system can not be transferred except by adding energy to it from without. The large values for $2a$ corresponding to large n need not, therefore, be contrary to experiment; indeed, we may in these large values seek an explanation of the fact, that in the laboratory it has hitherto not been possible to observe the hydrogen lines corresponding to large values of n in Balmer's formula, while they have been observed in the spectra of certain stars. In order that the large orbits of the electrons may not be disturbed by electrical forces from the neighbouring atoms the pressure will have to be very low, so low, indeed, that it is impossible to obtain sufficient light from a Geissler tube of ordinary dimensions. In the stars, however, we may assume that we have to do with hydrogen which is exceedingly attenuated and distributed throughout an enormously large region of space.

The Pickering lines. You have probably noticed that we have not mentioned at all the spectrum found in certain stars which according to the opinion then current was assigned to hydrogen, and together with the ordinary hydrogen spectrum was considered by Rydberg to form a connected system of lines completely analogous to the spectra of other elements. You have probably also perceived that difficulties would arise in interpreting this spectrum by means of the assumptions which have been employed. If such an attempt were to be made it would be necessary to give up the simple considerations which lead to the expression (7) for the constant R. We shall see, however, that it appears possible to explain the occurrence of this spectrum in another way. Let us suppose that it is not due to hydrogen, but to some other simple system consisting of a single electron revolving about a nucleus with an electrical charge Ne. The expression for ω becomes then

$$\omega^2 = \frac{2}{\pi^2} \frac{W^3}{N^2 e^4 m}.$$

Repeating the same calculations as before only in the inverse order we find, that this system will emit a line spectrum given by the expression

$$\frac{1}{\lambda} = \frac{2\pi^2 N^2 e^4 m}{ch^3} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) = R \left(\frac{1}{\left(\frac{n_1}{N}\right)^2} - \frac{1}{\left(\frac{n_2}{N}\right)^2}\right) \quad \text{......(9)}$$

By comparing this formula with the formula for Pickering's and Rydberg's series, we see that the observed lines can be explained on the basis of the theory, if it be assumed that the spectrum is due to an electron revolving about a nucleus with a charge $2e$, or according to Rutherford's theory around the nucleus of a helium atom. The fact that the spectrum in question is not observed in an ordinary helium tube, but only in stars, may be accounted for by the high degree of ionization which is required for the production of this spectrum; a neutral helium atom contains of course two electrons while the system under consideration contains only one.

These conclusions appear to be supported by experiment. Fowler, as I have mentioned, has recently succeeded in observing Pickering's and Rydberg's lines in a laboratory experiment. By passing a very heavy current through a mixture of hydrogen and helium Fowler observed not only these lines but also a new series of lines. This new series was of the same general type, the wave length being given approximately by

$$\frac{1}{\lambda} = R\left(\frac{1}{(\frac{3}{2})^2} - \frac{1}{(n+\frac{1}{2})^2}\right).$$

Fowler interpreted all the observed lines as the hydrogen spectrum sought for. With the observation of the latter series of lines, however, the basis of the analogy between the hypothetical hydrogen spectrum and the other spectra disappeared, and thereby also the foundation upon which Rydberg had founded his conclusions; on the contrary it is seen, that the occurrence of the lines was exactly what was to be expected on our view.

In the following table the first column contains the wave lengths measured by Fowler, while the second contains the limiting values of the experimental errors given by him; in the third column we find the products of the wave lengths by the quantity $\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) 10^{10}$; the values employed for n_1 and n_2 are enclosed in parentheses in the last column.

$\lambda . 10^8$	Limit of error	$\lambda . \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) . 10^{10}$	
4685·98	0·01	22779·1	(3 : 4)
3203·30	0·05	22779·0	(3 : 5)
2733·34	0·05	22777·8	(3 : 6)
2511·31	0·05	22778·3	(3 : 7)
2385·47	0·05	22777·9	(3 : 8)
2306·20	0·10	22777·3	(3 : 9)
2252·88	0·10	22779·1	(3 : 10)
5410·5	1·0	22774	(4 : 7)
4541·3	0·25	22777	(4 : 9)
4200·3	0·5	22781	(4 : 11)

The values of the products are seen to be very nearly equal, while the deviations are of the same order of magnitude as the limits of experimental error. The value of the product

$$\lambda\left(\frac{1}{n_1^2}-\frac{1}{n_2^2}\right)$$

should for this spectrum, according to the formula (9), be exactly ¼ of the corresponding product for the hydrogen spectrum. From the tables on pages 1 and 16 we find for these products 91153 and 22779, and dividing the former by the latter we get 4·0016. This value is very nearly equal to 4; the deviation is, however, much greater than can be accounted for in any way by the errors of the experiments. It has been easy, however, to find a theoretical explanation of this point. In all the foregoing calculations we have assumed that the mass of the nucleus is infinitely great compared to that of the electron. This is of course not the case, even though it holds to a very close approximation; for a hydrogen atom the ratio of the mass of the nucleus to that of the electron will be about 1850 and for a helium atom four times as great.

If we consider a system consisting of an electron revolving about a nucleus with a charge Ne and a mass M, we find the following expression for the frequency of revolution of the system:

$$\omega^2 = \frac{2}{\pi^2}\frac{W^3(M+m)}{N^2e^4Mm}.$$

From this formula we find in a manner quite similar to that previously employed that the system will emit a line spectrum, the wave lengths of which are given by the formula

$$\frac{1}{\lambda} = \frac{2\pi^2N^2e^4mM}{ch^3(M+m)}\left(\frac{1}{n_1^2}-\frac{1}{n_2^2}\right). \quad \text{..............(10)}$$

If with the aid of this formula we try to find the ratio of the product for the hydrogen spectrum to that of the hypothetical helium spectrum we get the value 4·00163 which is in complete agreement with the preceding value calculated from the experimental observations.

I must further mention that Evans has made some experiments to determine whether the spectrum in question is due to hydrogen or helium. He succeeded in observing one of the lines in very pure helium; there was, at any rate, not enough hydrogen present to enable the hydrogen lines to be observed. Since in any event Fowler does not seem to consider such evidence as conclusive it is to be hoped that these experiments will be continued. There is, however, also another possibility of deciding this question. As is evident from the formula (10), the helium spectrum under consideration should contain, besides the lines observed by Fowler, a series of lines lying close to the ordinary hydrogen lines. These lines may be obtained by putting $n_1 = 4$, $n_2 = 6, 8, 10$, etc. Even if these lines were present, it would be extremely difficult to observe them on account of their position with regard to the hydrogen lines, but should they be observed this would probably also settle the question of the origin of the spectrum, since no reason would seem to be left to assume the spectrum to be due to hydrogen.

Other spectra. For the spectra of other elements the problem becomes more complicated, since the atoms contain a larger number of electrons. It has not yet been possible on the basis of this theory to explain any other spectra besides those which I have already mentioned. On the other hand it ought to be mentioned that the general laws applying to the spectra are very simply interpreted on the basis of our assumptions. So far as the combination principle is concerned its explanation is obvious. In the method we have employed our point of departure was largely determined by this particular principle. But a simple explanation can be also given of the other general law, namely, the occurrence of Rydberg's constant in all spectral formulae. Let us assume that the spectra under consideration, like the spectrum of hydrogen, are emitted by a neutral system, and that they are produced by the binding of an electron previously removed from the system. If such an electron revolves about the nucleus in an orbit which is large in proportion to that of the other electrons it will be subjected to forces much the same as the electron in a hydrogen atom, since the inner electrons individually will approximately neutralize the effect of a part of the positive charge of the nucleus. We may therefore assume that for this system there will exist a series of stationary states in which the motion of the outermost electron is approximately the same as in the stationary states of a hydrogen atom. I shall not discuss these matters any further, but shall only mention that they lead to the conclusion that Rydberg's constant is not exactly the same for all elements. The expression for this constant will in fact contain the factor $\frac{M}{M+m}$, where M is the mass of the nucleus. The correction is exceedingly small for elements of large atomic weight, but for hydrogen it is, from the point of view of spectrum analysis, very considerable. If the procedure employed leads to correct results, it is not therefore permissible to calculate Rydberg's constant directly from the hydrogen spectrum; the value of the universal constant should according to the theory be 109735 and not 109675.

I shall not tire you any further with more details; I hope to return to these questions here in the Physical Society, and to show how, on the basis of the underlying ideas, it is possible to develop a theory for the structure of atoms and molecules. Before closing I only wish to say that I hope I have expressed myself sufficiently clearly so that you have appreciated the extent to which these considerations conflict with the admirably coherent group of conceptions which have been rightly termed the classical theory of electrodynamics. On the other hand, by emphasizing this conflict, I have tried to convey to you the impression that it may be also possible in the course of time to discover a certain coherence in the new ideas.

of lanthanum is 7/2, hence the nuclear magnetic moment as determined by this analysis is 2.5 nuclear magnetons. This is in fair agreement with the value 2.8 nuclear magnetons determined from La III hyperfine structures by the writer and N. S. Grace.[9]

[9] M. F. Crawford and N. S. Grace, Phys. Rev. **47**, 536 (1935).

This investigation was carried out under the supervision of Professor G. Breit, and I wish to thank him for the invaluable advice and assistance so freely given. I also take this opportunity to acknowledge the award of a Fellowship by the Royal Society of Canada, and to thank the University of Wisconsin and the Department of Physics for the privilege of working here.

MAY 15, 1935 PHYSICAL REVIEW VOLUME 47

Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?

A. EINSTEIN, B. PODOLSKY AND N. ROSEN, *Institute for Advanced Study, Princeton, New Jersey*

(Received March 25, 1935)

In a complete theory there is an element corresponding to each element of reality. A sufficient condition for the reality of a physical quantity is the possibility of predicting it with certainty, without disturbing the system. In quantum mechanics in the case of two physical quantities described by non-commuting operators, the knowledge of one precludes the knowledge of the other. Then either (1) the description of reality given by the wave function in quantum mechanics is not complete or (2) these two quantities cannot have simultaneous reality. Consideration of the problem of making predictions concerning a system on the basis of measurements made on another system that had previously interacted with it leads to the result that if (1) is false then (2) is also false. One is thus led to conclude that the description of reality as given by a wave function is not complete.

1.

ANY serious consideration of a physical theory must take into account the distinction between the objective reality, which is independent of any theory, and the physical concepts with which the theory operates. These concepts are intended to correspond with the objective reality, and by means of these concepts we picture this reality to ourselves.

In attempting to judge the success of a physical theory, we may ask ourselves two questions: (1) "Is the theory correct?" and (2) "Is the description given by the theory complete?" It is only in the case in which positive answers may be given to both of these questions, that the concepts of the theory may be said to be satisfactory. The correctness of the theory is judged by the degree of agreement between the conclusions of the theory and human experience. This experience, which alone enables us to make inferences about reality, in physics takes the form of experiment and measurement. It is the second question that we wish to consider here, as applied to quantum mechanics.

Whatever the meaning assigned to the term *complete*, the following requirement for a complete theory seems to be a necessary one: *every element of the physical reality must have a counterpart in the physical theory.* We shall call this the condition of completeness. The second question is thus easily answered, as soon as we are able to decide what are the elements of the physical reality.

The elements of the physical reality cannot be determined by *a priori* philosophical considerations, but must be found by an appeal to results of experiments and measurements. A comprehensive definition of reality is, however, unnecessary for our purpose. We shall be satisfied with the following criterion, which we regard as reasonable. *If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.* It seems to us that this criterion, while far from exhausting all possible ways of recognizing a physical reality, at least provides us with one such way, whenever the conditions set down in it occur. Regarded not as a necessary, but merely as a sufficient, condition of reality, this criterion is in agreement with classical as well as quantum-mechanical ideas of reality.

To illustrate the ideas involved let us consider the quantum-mechanical description of the behavior of a particle having a single degree of freedom. The fundamental concept of the theory is the concept of *state*, which is supposed to be completely characterized by the wave function ψ, which is a function of the variables chosen to describe the particle's behavior. Corresponding to each physically observable quantity A there is an operator, which may be designated by the same letter.

If ψ is an eigenfunction of the operator A, that is, if

$$\psi' \equiv A\psi = a\psi, \tag{1}$$

where a is a number, then the physical quantity A has with certainty the value a whenever the particle is in the state given by ψ. In accordance with our criterion of reality, for a particle in the state given by ψ for which Eq. (1) holds, there is an element of physical reality corresponding to the physical quantity A. Let, for example,

$$\psi = e^{(2\pi i/h)p_0 x}, \tag{2}$$

where h is Planck's constant, p_0 is some constant number, and x the independent variable. Since the operator corresponding to the momentum of the particle is

$$p = (h/2\pi i)\partial/\partial x, \tag{3}$$

we obtain

$$\psi' = p\psi = (h/2\pi i)\partial\psi/\partial x = p_0\psi. \tag{4}$$

Thus, in the state given by Eq. (2), the momentum has certainly the value p_0. It thus has meaning to say that the momentum of the particle in the state given by Eq. (2) is real.

On the other hand if Eq. (1) does not hold, we can no longer speak of the physical quantity A having a particular value. This is the case, for example, with the coordinate of the particle. The operator corresponding to it, say q, is the operator of multiplication by the independent variable. Thus,

$$q\psi = x\psi \neq a\psi. \tag{5}$$

In accordance with quantum mechanics we can only say that the relative probability that a measurement of the coordinate will give a result lying between a and b is

$$P(a, b) = \int_a^b \bar{\psi}\psi dx = \int_a^b dx = b - a. \tag{6}$$

Since this probability is independent of a, but depends only upon the difference $b-a$, we see that all values of the coordinate are equally probable.

A definite value of the coordinate, for a particle in the state given by Eq. (2), is thus not predictable, but may be obtained only by a direct measurement. Such a measurement however disturbs the particle and thus alters its state. After the coordinate is determined, the particle will no longer be in the state given by Eq. (2). The usual conclusion from this in quantum mechanics is that *when the momentum of a particle is known, its coordinate has no physical reality.*

More generally, it is shown in quantum mechanics that, if the operators corresponding to two physical quantities, say A and B, do not commute, that is, if $AB \neq BA$, then the precise knowledge of one of them precludes such a knowledge of the other. Furthermore, any attempt to determine the latter experimentally will alter the state of the system in such a way as to destroy the knowledge of the first.

From this follows that either (1) *the quantum-mechanical description of reality given by the wave function is not complete* or (2) *when the operators corresponding to two physical quantities do not commute the two quantities cannot have simultaneous reality.* For if both of them had simultaneous reality—and thus definite values—these values would enter into the complete description, according to the condition of completeness. If then the wave function provided such a complete description of reality, it would contain these values; these would then be predictable. This not being the case, we are left with the alternatives stated.

In quantum mechanics it is usually assumed that the wave function *does* contain a complete description of the physical reality of the system in the state to which it corresponds. At first sight this assumption is entirely reasonable, for the information obtainable from a wave function seems to correspond exactly to what can be measured without altering the state of the system. We shall show, however, that this assumption, together with the criterion of reality given above, leads to a contradiction.

2.

For this purpose let us suppose that we have two systems, I and II, which we permit to interact from the time $t=0$ to $t=T$, after which time we suppose that there is no longer any interaction between the two parts. We suppose further that the states of the two systems before $t=0$ were known. We can then calculate with the help of Schrödinger's equation the state of the combined system I+II at any subsequent time; in particular, for any $t>T$. Let us designate the corresponding wave function by Ψ. We cannot, however, calculate the state in which either one of the two systems is left after the interaction. This, according to quantum mechanics, can be done only with the help of further measurements, by a process known as the *reduction of the wave packet*. Let us consider the essentials of this process.

Let $a_1, a_2, a_3, \cdots$ be the eigenvalues of some physical quantity A pertaining to system I and $u_1(x_1), u_2(x_1), u_3(x_1), \cdots$ the corresponding eigenfunctions, where x_1 stands for the variables used to describe the first system. Then Ψ, considered as a function of x_1, can be expressed as

$$\Psi(x_1, x_2)=\sum_{n=1}^{\infty} \psi_n(x_2)u_n(x_1), \tag{7}$$

where x_2 stands for the variables used to describe the second system. Here $\psi_n(x_2)$ are to be regarded merely as the coefficients of the expansion of Ψ into a series of orthogonal functions $u_n(x_1)$. Suppose now that the quantity A is measured and it is found that it has the value a_k. It is then concluded that after the measurement the first system is left in the state given by the wave function $u_k(x_1)$, and that the second system is left in the state given by the wave function $\psi_k(x_2)$. This is the process of reduction of the wave packet; the wave packet given by the infinite series (7) is reduced to a single term $\psi_k(x_2)u_k(x_1)$.

The set of functions $u_n(x_1)$ is determined by the choice of the physical quantity A. If, instead of this, we had chosen another quantity, say B, having the eigenvalues $b_1, b_2, b_3, \cdots$ and eigenfunctions $v_1(x_1), v_2(x_1), v_3(x_1), \cdots$ we should have obtained, instead of Eq. (7), the expansion

$$\Psi(x_1, x_2)=\sum_{s=1}^{\infty} \varphi_s(x_2)v_s(x_1), \tag{8}$$

where φ_s's are the new coefficients. If now the quantity B is measured and is found to have the value b_r, we conclude that after the measurement the first system is left in the state given by $v_r(x_1)$ and the second system is left in the state given by $\varphi_r(x_2)$.

We see therefore that, as a consequence of two different measurements performed upon the first system, the second system may be left in states with two different wave functions. On the other hand, since at the time of measurement the two systems no longer interact, no real change can take place in the second system in consequence of anything that may be done to the first system. This is, of course, merely a statement of what is meant by the absence of an interaction between the two systems. Thus, *it is possible to assign two different wave functions* (in our example ψ_k and φ_r) *to the same reality* (the second system after the interaction with the first).

Now, it may happen that the two wave functions, ψ_k and φ_r, are eigenfunctions of two noncommuting operators corresponding to some physical quantities P and Q, respectively. That this may actually be the case can best be shown by an example. Let us suppose that the two systems are two particles, and that

$$\Psi(x_1, x_2)=\int_{-\infty}^{\infty} e^{(2\pi i/h)(x_1-x_2+x_0)p}dp, \tag{9}$$

where x_0 is some constant. Let A be the momentum of the first particle; then, as we have seen in Eq. (4), its eigenfunctions will be

$$u_p(x_1)=e^{(2\pi i/h)px_1} \tag{10}$$

corresponding to the eigenvalue p. Since we have here the case of a continuous spectrum, Eq. (7) will now be written

$$\Psi(x_1, x_2)=\int_{-\infty}^{\infty}\psi_p(x_2)u_p(x_1)dp, \qquad (11)$$

where

$$\psi_p(x_2)=e^{-(2\pi i/h)(x_2-x_0)p}. \qquad (12)$$

This ψ_p however is the eigenfunction of the operator

$$P=(h/2\pi i)\partial/\partial x_2, \qquad (13)$$

corresponding to the eigenvalue $-p$ of the momentum of the second particle. On the other hand, if B is the coordinate of the first particle, it has for eigenfunctions

$$v_x(x_1)=\delta(x_1-x), \qquad (14)$$

corresponding to the eigenvalue x, where $\delta(x_1-x)$ is the well-known Dirac delta-function. Eq. (8) in this case becomes

$$\Psi(x_1, x_2)=\int_{-\infty}^{\infty}\varphi_x(x_2)v_x(x_1)dx, \qquad (15)$$

where

$$\varphi_x(x_2)=\int_{-\infty}^{\infty}e^{(2\pi i/h)(x-x_2+x_0)p}dp$$
$$=h\delta(x-x_2+x_0). \qquad (16)$$

This φ_x, however, is the eigenfunction of the operator

$$Q=x_2 \qquad (17)$$

corresponding to the eigenvalue $x+x_0$ of the coordinate of the second particle. Since

$$PQ-QP=h/2\pi i, \qquad (18)$$

we have shown that it is in general possible for ψ_k and φ_r to be eigenfunctions of two noncommuting operators, corresponding to physical quantities.

Returning now to the general case contemplated in Eqs. (7) and (8), we assume that ψ_k and φ_r are indeed eigenfunctions of some noncommuting operators P and Q, corresponding to the eigenvalues p_k and q_r, respectively. Thus, by measuring either A or B we are in a position to predict with certainty, and without in any way disturbing the second system, either the value of the quantity P (that is p_k) or the value of the quantity Q (that is q_r). In accordance with our criterion of reality, in the first case we must consider the quantity P as being an element of reality, in the second case the quantity Q is an element of reality. But, as we have seen, both wave functions ψ_k and φ_r belong to the same reality.

Previously we proved that either (1) the quantum-mechanical description of reality given by the wave function is not complete or (2) when the operators corresponding to two physical quantities do not commute the two quantities cannot have simultaneous reality. Starting then with the assumption that the wave function does give a complete description of the physical reality, we arrived at the conclusion that two physical quantities, with noncommuting operators, can have simultaneous reality. Thus the negation of (1) leads to the negation of the only other alternative (2). We are thus forced to conclude that the quantum-mechanical description of physical reality given by wave functions is not complete.

One could object to this conclusion on the grounds that our criterion of reality is not sufficiently restrictive. Indeed, one would not arrive at our conclusion if one insisted that two or more physical quantities can be regarded as simultaneous elements of reality *only when they can be simultaneously measured or predicted.* On this point of view, since either one or the other, but not both simultaneously, of the quantities P and Q can be predicted, they are not simultaneously real. This makes the reality of P and Q depend upon the process of measurement carried out on the first system, which does not disturb the second system in any way. No reasonable definition of reality could be expected to permit this.

While we have thus shown that the wave function does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible.

I

THE UNDULATORY ASPECTS OF THE ELECTRON[1]

WHEN, in 1920, I resumed my investigations in theoretical Physics after a long interruption through circumstances out of my own control, I was far from imagining that this research would within a few years be rewarded by the lofty and coveted distinction given each year by the Swedish Academy of Sciences: the Nobel Prize in Physics. At that time what drew me towards theoretical Physics was not the hope that so high a distinction would ever crown my labours: what attracted me was the mystery which was coming to envelop more and more deeply the structure of Matter and of radiation in proportion as the strange concept of the quantum, introduced by Planck about 1900 during his research on black body radiation, came to extend over the entire field of Physics.

But to explain the way in which my research came to develop I must first outline the critical period through which Physics had for the last twenty years been passing.

* * *

Physicists had for long been wondering whether Light did not consist of minute corpuscles in rapid motion, an idea going back to the philosophers of antiquity, and sustained in the eighteenth century by Newton. After interference phenomena had been discovered by Thomas Young, however, and Augustin Fresnel had completed his important investigation, the assumption that Light had a granular structure was entirely disregarded, and the

[1] Address delivered at Stockholm on receiving the Nobel Prize, December 12, 1929.

Wave Theory was unanimously adopted. In this way the physicists of last century came to abandon completely the idea that Light had an atomic structure. But the Atomic Theory, being thus banished from optics, began to achieve great success, not only in Chemistry, where it provided a simple explanation of the laws of definite proportions, but also in pure Physics, where it enabled a fair number of the properties of solids, liquids and gases to be interpreted. Among other things it allowed the great kinetic theory of gases to be formulated, which, in the generalized form of statistical Mechanics, has enabled clear significance to be given to the abstract concepts of thermodynamics. We have seen how decisive evidence in favour of the atomic structure of electricity was also provided by experiments. Thanks to Sir J. J. Thomson, the notion of the corpuscle of electricity was introduced; and the way in which H. A. Lorentz has exploited this idea in his electron Theory is well known.

Some thirty years ago, then, Physics was divided into two camps. On the one hand there was the Physics of Matter, based on the concepts of corpuscles and atoms which were assumed to obey the classical laws of Newtonian Mechanics; on the other hand there was the Physics of radiation, based on the idea of wave propagation in a hypothetical continuous medium: the ether of Light and of electromagnetism. But these two systems of Physics could not remain alien to each other: an amalgamation had to be effected; and this was done by means of a theory of the exchange of energy between Matter and radiation. It was at this point, however, that the difficulties began; for in the attempt to render the two systems of Physics compatible with each other, incorrect and even impossible conclusions were reached with regard to the energy equilibrium between Matter and radiation in an enclosed and thermally isolated region: some investigators even going so far as to say that Matter would transfer all its energy to radiation, and hence tend towards the temperature of absolute zero. This absurd conclusion had to be avoided at all costs; and by a brilliant piece of intuition Planck succeeded in doing so. Instead of assuming, as did the classical Wave Theory, that a light-source emits its radiation continuously, he assumed that it emits it in equal and finite quantities—in quanta. The energy of each quantum, still further, was supposed to be proportional to the frequency of the radiation, ν, and to be equal to $h\nu$, where h is the universal constant since known as Planck's Constant.

The success of Planck's ideas brought with it some serious consequences. For if Light is emitted in quanta, then surely, once radiated, it ought to have a granular structure. Consequently the existence of quanta of radiation brings us back to the corpuscular conception of Light. On the other hand, it can be shown—as has in fact been done by Jeans and H. Poincaré—that if the motion of the material particles in a light-source obeyed the laws of classical Mechanics, we could never obtain the correct Law of black body radiation—Planck's Law. It must therefore be admitted that the older dynamics, even as modified by Einstein's Theory of Relativity, cannot explain motion on a very minute scale.

The existence of a corpuscular structure of Light and of other types of radiation has been confirmed by the discovery of the photo-electric effect which, as I have already observed, is easily explained by the assumption that the radiation consists of quanta—$h\nu$—capable of transferring their entire energy to an electron in the irradiated substance; and in this way we are brought to the theory of light-quanta which, as we have seen, was advanced in 1905 by Einstein—a theory which amounts to a return to Newton's corpuscular hypothesis, supplemented by the proportionality subsisting between the energy of the corpuscles and the frequency. A number of arguments were adduced by Einstein in support of his view, which was confirmed by Compton's discovery in 1922 of the scattering of X-rays, a phenomenon named after him. At the same time it still remained necessary to retain the Wave Theory to explain the phenomena of diffraction and interference, and no means was apparent to reconcile this Theory with the existence of light-corpuscles.

I have pointed out that in the course of investigation some doubt had been thrown on the validity of small-scale Mechanics. Let us imagine a material point describing a small closed orbit—

an orbit returning on itself; then according to classical dynamics there is an infinity of possible movements of this type in accordance with the initial conditions, and the possible values of the energy of the moving material point form a continuous series. Planck, on the other hand, was compelled to assume that only certain privileged movements—*quantized* motion—are possible, or at any rate stable, so that the available values of the energy form a discontinuous series. At first this seemed a very strange idea; soon, however, its truth had to be admitted, because it was by its means that Planck arrived at the correct Law of black body radiation and because its usefulness has since been proved in many other spheres. Finally, Bohr founded his famous atomic Theory on this idea of the quantization of atomic motion—a theory so familiar to scientists that I will refrain from summing it up here.

Thus we see once again it had become necessary to assume two contradictory theories of Light, in terms of waves, and of corpuscles, respectively; while it was impossible to understand why, among the infinite number of paths which an electron ought to be able to follow in the atom according to classical ideas, there was only a restricted number which it could pursue in fact. Such were the problems facing physicists at the time when I returned to my studies.

* * *

When I began to consider these difficulties I was chiefly struck by two facts. On the one hand the Quantum Theory of Light cannot be considered satisfactory, since it defines the energy of a light-corpuscle by the equation $W = h\nu$, containing the frequency ν. Now a purely corpuscular theory contains nothing that enables us to define a frequency; for this reason alone, therefore, we are compelled, in the case of Light, to introduce the idea of a corpuscle and that of periodicity simultaneously.

On the other hand, determination of the stable motion of electrons in the atom introduces integers; and up to this point the only phenomena involving integers in Physics were those of interference and of normal modes of vibration. This fact suggested to me the idea that electrons too could not be regarded simply as corpuscles, but that periodicity must be assigned to them also.

In this way, then, I obtained the following general idea, in accordance with which I pursued my investigations:—that it is necessary in the case of Matter, as well as of radiation generally and of Light in particular, to introduce the idea of the corpuscle and of the wave simultaneously: or in other words, in the one case as well as in the other, we must assume the existence of corpuscles accompanied by waves. But corpuscles and waves cannot be independent of each other: in Bohr's terms, they are two complementary aspects of Reality: and it must consequently be possible to establish a certain parallelism between the motion of a corpuscle and the propagation of its associated wave. The first object at which to aim, therefore, was to establish the existence of this parallelism.

With this in view, I began by considering the simplest case: that of an isolated corpuscle, i.e. one removed from all external influence; with this we wish to associate a wave. Let us therefore consider first of all a reference system $O\,x_0 y_0 z_0$ in which the corpuscle is at rest: this is the "proper" system for the corpuscle according to the Theory of Relativity. Within such a system the wave will be stationary, since the corpuscle is at rest; its phase will be the same at every point, and it will be represented by an expression of the form $\sin 2\pi\nu_0(t_0 - \tau_0)$, t_0 being the "proper" time of the corpuscle, and τ_0 a constant.

According to the principle of inertia the corpuscle will be in uniform rectilinear motion in every Galilean system. Let us consider such a Galilean system, and let v be the velocity of the corpuscle in this system. Without loss of generality, we may take the direction of motion to be the axis of x. According to the Lorentz transformation, the time t employed by an observer in this new system is linked with the proper time t_0 by the relation:

$$t_0 = \frac{t - \dfrac{\beta x}{c}}{\sqrt{1 - \beta^2}}$$

where $\beta = v/c$.

Hence for such an observer the phase of the wave will be given by

$$\sin 2\pi \frac{\nu_0}{\sqrt{1-\beta^2}}\left(t - \frac{\beta x}{c} - \tau_0\right).$$

Consequently the wave will have for him a frequency

$$\nu = \frac{\nu_0}{\sqrt{1-\beta^2}}$$

and will move along the axis of x with the phase-velocity

$$V = \frac{c}{\beta} = \frac{c^2}{v}.$$

If we eliminate β from the two preceding formulae we shall easily find the following relation, which gives the index of refraction of free space, n, for the waves under consideration

$$n = \sqrt{1 - \frac{\nu_0^2}{\nu^2}}$$

To this "law of dispersion" there corresponds a "group velocity." You are aware that the group velocity is the velocity with which the resultant amplitude of a group of waves, with almost equal frequencies, is propagated. Lord Rayleigh has shown that this velocity U satisfies the equation

$$\frac{1}{U} = \frac{1}{c}\frac{d(n\nu)}{d\nu}$$

Here we find that $U = v$, which means that the velocity of the group of waves in the system $xyzt$ is equal to the velocity of the corpuscle in this system. This relation is of the greatest importance for the development of the Theory.

Accordingly, in the system $xyzt$ the corpuscle is defined by the frequency ν and by the phase-velocity V of its associated wave. In order to establish the parallelism mentioned above, we must try to connect these magnitudes to the mechanical magnitudes—to energy and momentum. The ratio between energy and frequency is one of the most characteristic relations of the Quantum Theory; and since, still further, energy and frequency are transformed when the Galilean system of reference is changed, it is natural to establish the equation

$$\text{Energy} = h \times \text{frequency, or } W = h\nu$$

where h is Planck's constant. This relation must apply to all Galilean systems; and in the proper system of the corpuscle where, according to Einstein, the energy of the corpuscle is reduced to its internal energy m_0c^2 (where m_0 is its proper mass) we have $h\nu_0 = m_0c^2$.

This relation gives the frequency ν_0 as a function of the proper mass m_0, or inversely.

The momentum is a vector $\mathbf{p}$ equal to $\frac{m_0\mathbf{v}}{\sqrt{1-\beta^2}}$, where $|\mathbf{v}| = v$, and then we have

$$p = |\mathbf{p}| = \frac{m_0 v}{\sqrt{1-\beta^2}} = \frac{Wv}{c^2} = \frac{h\nu}{V} = \frac{h}{\lambda}.$$

The quantity λ is the wave-length—the distance between two consecutive wave-crests

hence $$\lambda = \frac{h}{p}.$$

This is a fundamental relation of the Theory.

* * *

All that has been said refers to the very simple case where there is no field of force acting on the corpuscle. I shall now indicate very briefly how the Theory can be generalized for the case of a corpuscle moving in a field of force not varying with time derived from a potential energy function $F(x, y, z)$. Arguments into which I shall not enter here lead us in such a case to assume that the propagation of the wave corresponds to an index of refraction varying from point to point in space in accordance with the formula

$$n(x, y, z) = \sqrt{\left[1 - \frac{F(x,y,z)}{h\nu}\right]^2 - \frac{\nu_0^2}{\nu^2}}$$

or, as a first approximation, if we neglect the corrections introduced by the Theory of Relativity

$$n(x, y, z) = \sqrt{\frac{2(E - F)}{m_0 c^2}} \quad \text{with} \quad E = W - m_0 c^2.$$

The constant energy W of the corpuscle is further connected with the constant frequency ν of the wave by the relation

$$W = h\nu$$

while the wave-length λ, which varies from one point to the other in the field of force, is connected with the momentum p (which is also variable) by the relation

$$\lambda(x, y, z) = \frac{h}{p(x, y, z)}.$$

Here again we show that the velocity of the wave-group is equal to the velocity of the corpuscle. The parallelism thus established between the corpuscle and its wave enables us to identify Fermat's Principle in the case of waves and the Principle of Least Action in that of corpuscles, for constant fields. Fermat's Principle states that the ray in the optical sense passing between two points A and B in a medium whose index is $n(x, y, z)$, variable from one point to the other but constant in time, is such that the integral $\int_A^B n\, dl$, taken along this ray, shall be an *extremum*. On the other hand, Maupertuis' Principle of Least Action asserts that the trajectory of a corpuscle passing through two points A and B is such that the integral $\int_A^B p\, dl$ taken along the trajectory shall be an *extremum*, it being understood that we are considering only the motion corresponding to a given value of energy. According to the relations already established between the mechanical and the wave magnitudes, we have

$$n = \frac{c}{V} = \frac{c}{\nu} \cdot \frac{1}{\lambda} = \frac{c}{h\nu} \cdot \frac{h}{\lambda} = \frac{c}{W} p = \text{constant} \times p$$

since W is constant in a constant field. Hence it follows that Fermat's Principle and Maupertuis' Principle are each a rendering of the other: the possible trajectories of the corpuscle are identical with the possible rays of its wave.

These ideas lead to an interpretation of the conditions of stability introduced by the Quantum Theory. If we consider a closed trajectory C a constant field it is quite natural to assume that the phase of the associated wave should be a uniform function along this trajectory. This leads us to write

$$\int_C \frac{dl}{\lambda} = \int_C \frac{1}{h} p \, dl = \text{an integer.}$$

Now this is exactly the condition of the stability of atomic periodic motion, according to Planck. Thus the quantum conditions of stability appear as analogous to resonance phenomena, and the appearance of integers here becomes as natural as in the theory of vibrating cords and discs.

* * *

The general formulae establishing the parallelism between waves and corpuscles can be applied to light-corpuscles if we assume that in that case the rest-mass m_0 is infinitely small. If then for any given value of the energy W we make m_0 tend to zero, we find that both v and V tend to c, and in the limit we obtain the two fundamental formulae on which Einstein erected his Theory of Light-quanta

$$W = h\nu \quad p = \frac{h\nu}{c}.$$

Such were the principal ideas which I had developed during my earlier researches. They showed clearly that it was possible to establish a correspondence between waves and corpuscles of such a kind that the Laws of Mechanics correspond to those of geometrical optics. But we know that in the Wave Theory geometrical optics is only an approximation: there are limits to the validity of this approximation, and especially when the phenomena of interference and of diffraction are concerned it is wholly inadequate. This suggests the idea that the older Mechanics too may be no more than an approximation as compared with a more comprehensive Mechanics of an undulatory character. This was what I expressed at the beginning of my researches when I said that a new Mechanics must be formulated, standing in the same relation to the older Mechanics as that in which wave optics stands to geometrical optics. This new Mechanics has since been developed, thanks in particular to the fine work done by Schrödinger. It starts from the equations of wave propagation, which are taken as the basis, and rigorously determines the temporal changes of the wave associated with a corpuscle. More particularly, it has succeeded in giving a new and more satisfactory form to the conditions governing the quantization of intra-atomic motion: for, as we have seen, the older conditions of quantization are encountered again if we apply geometrical optics to the waves associated with intra-atomic corpuscles; and there is strictly no justification for this application.

I cannot here trace even briefly the development of the new Mechanics. All that I wish to say is that on examination it has shown itself to be identical with a Mechanics developed independently, first by Heisenberg and later by Born, Jordan, Pauli, Dirac and others. This latter Mechanics—Quantum Mechanics—and Wave Mechanics are, from the mathematical point of view, equivalent to each other.

Here we must confine ourselves to a general consideration of the results obtained. To sum up the significance of Wave Mechanics, we can say that a wave must be associated with each particle, and that a study of the propagation of the wave alone can tell us anything about the successive localizations of the corpuscle in space. In the usual large-scale mechanical phenomena, the localizations predicted lie along a curve which is the trajectory in the classical sense of the term. What, however, happens if the wave is not propagated according to the laws of geometrical optics; if, for example, interference or diffraction occurs? In such a case we can no longer assign to the corpuscle motion in accordance with classical dynamics. So much is certain. But a further question arises: Can we even suppose that at any given moment the corpuscle has an exactly determined position within the wave, and that in the course of its propagation the wave carries the corpuscle with it, as a wave of water would carry a cork? These are difficult questions, and their discussion would carry us too far and actually to the borderland of Philosophy. All that I shall say here is that the general modern tendency is to assume that it is not always possible to assign an exactly defined position within the wave to the corpuscle, that whenever an observation is made enabling us to localize the corpuscle, we are invariably led to attribute to it a position inside the wave, and that the probability that this position is at a given point, M, within the wave is proportional to the square of the amplitude, or the intensity, at M.

What has just been said can also be expressed in the following way. If we take a cloud of corpuscles all associated with the same wave, then the intensity of the wave at any given point is proportional to the density of the cloud of corpuscles at that point, i.e. to the number of corpuscles per unit of volume around that point. This assumption must be made in order to explain how it is that in the case of interference the luminous energy is found concentrated at those points where the intensity of the wave is at a maximum: if it is assumed that the luminous energy is transferred by light-corpuscles, or photons, then it follows that the density of the photons in the wave is proportional to this intensity.

This rule by itself enables us to understand the way in which the undulatory theory of the electron has been verified experimentally.

For let us imagine an indefinite cloud of electrons, all having the same velocity and moving in the same direction. According to the fundamental ideas of Wave Mechanics, we must associate with this cloud an infinite plane wave having the form

$$a \text{ exp. } 2\pi i \left[\frac{W}{h} t - \frac{\alpha x + \beta y + \gamma z}{\lambda} \right]$$

where α, β, γ are the direction cosines of the direction of propagation, and where the wave-length λ is equal to $\frac{h}{p}$. If the electrons have no extremely high velocity, we may say

$$p = m_0 v$$

and hence

$$\lambda = \frac{h}{m_0 v}$$

m_0 being the rest-mass of the electron.

In practice, to obtain electrons having the same velocity they are subjected to the same potential difference P. We then have

$$\frac{1}{2} m_0 v^2 = e\,P$$

Consequently

$$\lambda = \frac{h}{\sqrt{2 m_0\, e\, P}}$$

Numerically, this gives

$$\lambda = \frac{12 \cdot 24}{\sqrt{P}} \cdot 10^{-8} \text{ cm. } (P \text{ in volts}).$$

As we can only use electrons that have fallen through a potential difference of at least some tens of volts, it follows that the wave-length λ, assumed by the Theory, is at most of the order of 10^{-8} cm., i.e. of the order of the Ångström unit.[1] This is also the order of magnitude of the wave-lengths of X-rays.

The length of the electron wave being thus of the same order as that of X-rays, we may fairly expect to be able to obtain a scattering of this wave by crystals, in complete analogy to the Laue phenomenon in which, in a natural crystal like rock salt, the atoms of the substances composing the crystal are arranged at regular intervals of the order of one Ångström, and thus act as scattering centres for the waves. If a wave having a length of one Ångström encounters the crystal, then the waves scattered agree

[1] [This is one ten-millionth mm.]

in phase in certain definite directions. In these directions the total intensity scattered exhibits a strong maximum. The location of these maxima of scattering is given by the well known mathematical Theory elaborated by Laue and Bragg, which gives the position of the maxima in terms of the distance between the atomic arrangements in the crystal and of the length of the incident wave. For X-rays the Theory has been triumphantly substantiated by Laue, Friedrich and Knipping, and today the diffraction of X-rays by crystals has become a quite commonplace experiment. The exact measurement of the wave-lengths of X-rays is based on this diffraction, as I need hardly recall in a country where Siegbahn and his collaborators are pursuing their successful labours.

In the case of X-rays, the phenomenon of diffraction by crystals was a natural consequence of the idea that these rays are undulations analogous to Light, and differ from Light only by their shorter wave-length. But for electrons no such view could be entertained, so long as the latter were looked upon as being merely minute corpuscles. If, on the other hand, we assume that the electron is associated with a wave, and that the density of a cloud of electrons is measured by the intensity of the associated wave, we may then expect that there will be effects in the case of electrons similar to the Laue effect. In that event, the electron wave will be scattered with an intensity in certain directions which the Laue-Bragg Theory enables us to calculate, on the assumption that the wave-length is $\lambda = \frac{h}{mv}$, a length corresponding to the known velocity v of the electrons falling on the crystal. According to our general principle, the intensity of the scattered wave measures the density of the cloud of scattered electrons, so that we may expect to find large numbers of scattered electrons in the directions of the maxima. If this effect actually occurs, it would provide a crucial experimental proof of the existence of a wave associated with the electron, its length being $\frac{h}{mv}$. In this way the fundamental idea of Wave Mechanics would be provided with a firm experimental foundation.

Now experiment—which is the last Court of Appeal of theories

—has shown that the diffraction of electrons by crystals actually occurs, and that it follows the Laws of Wave Mechanics exactly and quantitatively. It is (as we have seen already) to Davisson and Germer, working at the Bell Laboratories in New York, that the credit belongs of having been the first to observe this phenomenon by a method similar to that used by Laue for X-rays. Following up the same experiments, but substituting for the single crystal a crystalline powder, in accordance with the method introduced for X-rays by Debye and Scherrer, Professor G. P. Thomson, of Aberdeen, the son of the great Cambridge physicist, Sir J. J. Thomson, has discovered the same phenomena. At a later stage Rupp in Germany, Kikuchi in Japan and Ponte in France have also reproduced them under varying experimental conditions. Today the existence of the effect is no longer subject to doubt, and the minor difficulties of interpretation which Davisson's and Germer's earlier experiments had raised have been resolved in a satisfactory manner. Rupp has actually succeeded in obtaining the diffraction of electrons in a particularly striking form. A grating is employed—a metal or glass surface, either plane or slightly curved, on which equidistant lines have been mechanically drawn, the interval between them being of an order of magnitude comparable to that of the wave-lengths of Light. Between the waves diffracted by these lines there will be interference, and the interference will give rise to maxima of diffracted Light in certain directions depending on the distance between the lines, on the direction of the Light falling on the grating and on the wave-lengths. For a long time it remained impossible to obtain similar effects with gratings of this kind produced by human workmanship when X-rays were used instead of Light. The reason for this was that the wave-length of X-rays is a great deal shorter than that of Light, and that there is no instrument capable of drawing lines on any surface at intervals of the order of X-ray wave-lengths. Ingenious physicists, however, (Compton and Thibaud) succeeded in overcoming the difficulty. Let us take an ordinary optical grating and let us look at it more or less at a tangent. The lines of the grating will then seem to be much closer together than they actually are. For X-rays falling on the grating at this grazing angle, the conditions will be the same as though the lines were extremely close together, and diffraction effects like those of Light will be produced. The physicists just mentioned have proved that such was in fact the case. But now—since the electron wave-lengths are of the same order as those of X-rays—we should also be able to obtain these diffraction phenomena by causing a beam of electrons to fall on such an optical grating at a very small grazing angle. Rupp succeeded in doing this. He was thus enabled to measure the length of electron waves by comparing it directly with the distance between the lines drawn mechanically on the grating.

* * *

We thus find that in order to describe the properties of Matter, as well as those of Light, we must employ waves and corpuscles simultaneously. We can no longer imagine the electron as being just a minute corpuscle of electricity: we must associate a wave with it. And this wave is not just a fiction: its length can be measured and its interferences calculated in advance. In fact, a whole group of phenomena was in this way predicted before being actually discovered. It is, therefore, on this idea of the dualism in Nature between waves and corpuscles, expressed in a more or less abstract form, that the entire recent development of theoretical Physics has been built up, and that its immediate future development appears likely to be erected.